Learning from the review of "Estimating stock status from relative abundance and resilience"*

Chapter *in* Fisheries · March 2020

CITATIONS
0

READS
105

11 authors, including:

Henning Winker
European Commission
120 PUBLICATIONS **1,254** CITATIONS
SEE PROFILE

Gianpaolo Coro
Italian National Research Council
70 PUBLICATIONS **381** CITATIONS
SEE PROFILE

Nazli Demirel
Istanbul University
38 PUBLICATIONS **329** CITATIONS
SEE PROFILE

Athanassios C Tsikliras
Aristotle University of Thessaloniki
94 PUBLICATIONS **2,050** CITATIONS
SEE PROFILE

ISSN 1198-6727

Fisheries Centre Research Reports

2020 Volume 28 Number 2

Marine and Freshwater Miscellanea II

Institute for the Oceans and Fisheries,
The University of British Columbia, Canada

LEARNING FROM THE REVIEW OF "ESTIMATING STOCK STATUS FROM RELATIVE ABUNDANCE AND RESILIENCE"*

Rainer Froese[1], Henning Winker[2,3], Gianpaolo Coro[4], Nazli Demirel[5], Athanassios C. Tsikliras[6], Donna Dimarchopoulou[6], Giuseppe Scarcella[7], Maria Lourdes Deng Palomares[8], Manuel Dureuil[9,10], and Daniel Pauly[8]

1) *GEOMAR Helmholtz Centre for Ocean Research, Düsternbrooker Weg 20, 24105 Kiel, Germany*
2) *DEFF - Department of Environment, Forestry and Fisheries, Private Bag X2, Vlaeberg 8018, South Africa*
3) *Centre for Statistics in Ecology, Environment and Conservation, Department of Statistical Sciences, University of Cape Town, Private Bag X3 Rondebosch, South Africa*
4) *Institute of Information Science and Technologies "A. Faedo" - National Research Council of Italy (ISTI-CNR), via Moruzzi 1, 56124 Pisa, Italy*
5) *Institute of Marine Sciences and Management, Istanbul University, Istanbul 34134, Turkey*
6) *Laboratory of Ichthyology, School of Biology, Aristotle University of Thessaloniki, 54124 Thessaloniki, Greece*
7) *Institute of Marine Science - National Research Council of Italy (ISMAR-CNR), L.go Fiera della Pesca, 60125 Ancona, Italy*
8) *Sea Around Us, Institute for the Ocean and Fisheries, University of British Columbia, 2202 Main Mall, Vancouver, BC, V6T 1Z4, Canada*
9) *Department of Biology, Dalhousie University, 1355 Oxford St., PO BOX 15000, Halifax, NS, B3H 4R2, Canada*
10) *Sharks of the Atlantic Research and Conservation Centre, Halifax, NS, B3L 2Y5, Canada*

Abstract

This contribution presents the detailed responses to the peer-review of Froese *et al.* (2019) "Estimating stock status from relative abundance and resilience" (ICES J. Mar. Sci. 2019) which outlined a method called "AMSY" for inferring biomass trends for stocks for which only catch-per-unit-effort and limited ancillary ('priors') data are available. The responses emphasize that the required priors are legitimate and straightforward to obtain, thus, making AMSY a method of choice in data-sparse situations. This is also a good example of the role of peer-review in validating and improving science.

Introduction

Transparency is the lifeblood of science (Dittert *et al.* 2001) and various approaches have been and continue to be proposed to make more of the data used in scientific research widely available, e.g., in the marine sciences (Froese et al. 2001; Froese and Reyes 2003; Zeller et al. 2005).

Moreover, we believe that more of the back-and-forth dialogs that undergird scientific articles should also be made publicly available. Here, we follow Pauly (2018) in publishing the peer-review (i.e., the editors' and reviewers' comments and our responses) of our contribution entitled "Estimating stock status from relative abundance and resilience" (Froese et al. 2019), which presented a method called "AMSY" to derive a time series of biomass from catch-per-unit-effort (or CPUE) data and priors, notably for the growth rate of the population (resilience).

* Cite as: Froese, D., H. Winker, G. Coro, N. Demirel, A.C. Tsikliras, D. Dimarchopoulou, G. Scarcella, M.L.D. Palomares, M. Dureuil and D. Pauly. 2020. Learning from the peer review of 'Estimating stock status from relative abundance and resilience', p. 111-124. *In:* Pauly, D. and V. Ruiz-Leotaud (eds.) *Marine and Freshwater Miscellanea II*. Fisheries Centre Research Reports 28(2). Institute for the Oceans and Fisheries, University of British Columbia.

Editors' comments:

E1) The large number of priors and constraints give no confidence that the method is an improvement over looking at the CPUE trend and concluding the stock status.

E1. Answer: The number of priors (2) is actually low. The number of filters, which are not priors and need no user interaction, has been reduced from 7 to 5. The influence of these filters has been documented in a new section, by running the simulations with and without filters.

E2) The text does not adequately describe how the results are sensitive to prior specifications and the chosen set of filters based on the CPUE time-series.

E2. Answer: The text has been greatly expanded (see highlighted sections) to address these issues.

E3) There are many places where more work should be done, and more considerations should be made. There are many assumptions associated with this approach, only some of which are discussed, and none that are performance tested.

E3) Answer: All assumptions are now listed and discussed. A preliminary performance testing is done against 140 real world stocks. Still, it is stressed that more sensitivity studies can be done, but that this is beyond the scope of an already substantial study that presents a new method and proof of concept.

E4) There are many assumptions being made by this model; violating these assumptions and by how much to degrade performance is the true test of how this method performs, and this needs to be done much better in the paper.

E4. Answer: The assumptions made for AMSY are the same as for a regular Bayesian implementation of a surplus-production model, the only difference being the lack of catch data. This is now made explicit in the text. Also, the effect of the filters is now demonstrated by analyzing the simulated data with and without filters.

E5) There needs to be further description and reporting on the simulations from which filters are derived in order for the reader to understand the derivation of the filters.

E5. Answer: This presumably refers to the numerical settings/thresholds used for the filters. Their derivation is now described and declared as preliminary and open to further improvements, but assumed sufficient for the purpose of this proof-of-concept study.

E6) The supplementary materials (>200 pages!), as currently presented, will be overwhelming to most readers without further clarification.

E6. Answer: The amended/expanded text and the addition of a new table and a new figure make it less necessary to consult the Supplementary Materials. But properly documenting the results of 2*24 runs against simulated data, runs against 140 real stocks, and first assessments of 38 data-poor stocks does require a lot of space. We have added a more detailed table of contents to the appendices.

Reviewer 1 comments:

R1.1: I find no technical errors in the model implemented by the authors.

R1.1 Answer: We thank R1 for the technical review of our model and note that no errors were found.

R1.2: But the large number of priors and constraints give me no confidence that the method is an improvement over looking at the CPUE trend and concluding whether the stock seems to be going up or going down or seems fairly stable. Below are some particular comments.

R1.2 Answer: The number of priors is actually not large: one prior for productivity, objectively obtained from online databases based on life history traits or previous assessments, and one prior for relative stock size in any year of the time series, obtained preferably from independent data such as length frequencies, or from expert knowledge, such as is common practice in Bayesian modelling. If Reviewer 1 refers to the applied filters, their role is now explicitly clarified and their number has been reduced from 7 to 5.

R1.3: Line 126: relative stock size Bt/k: putting a prior on the result is assuming you know the answer.

R1.3 Answer: Putting priors on parameter estimates is required in Bayesian-like analyses, such as the proposed AMSY approach. Reviewer 1 reports a common criticism of Bayesian inference, which is, however, a rigorous state-of-the-art statistical procedure. Properly applied, priors reflect the best available knowledge about a trait or parameter, and that knowledge is then updated with the available data. The point here is whether a model that only has a time series of CPUE as input can produce similar results as a model that, in addition, has a time series of catch data as input, everything else being equal. This has now been stressed in the text.

R1.4: Also, good to acknowledge the work of Alec MacCall in this regard.

R1.4 Answer: Reviewer 1 seems to refer to the DCAC method of MacCall (2009), which estimates a sustainable catch-level below MSY and requires catch, relative depletion, M and FMSY/M estimates as inputs. Note the requirement for prior knowledge of relative depletion, which is the same as the B/k prior required by AMSY. DCAC needs a time series of catch data and thus is not applicable for the CPUE-only situations that AMSY is meant to deal with. Note the requirement for prior knowledge of M, which is similar to resilience: M ~ FMSY ~ 0.5 r. In other words, the example given by Reviewer 1 shows the common use of priors in stock assessment.

R1.5: Line 140-160: These are highly informative priors, so there is no surprise that the method works. If data-rich "full stock assessments" used such priors, they would perform exceedingly well.

R1.5 Answer: Bayesian age-structured assessments actually use the same priors, plus additional ones for natural mortality and the steepness of the left, ascending side of stock recruitment curves. While meta-analyses and largely reliable methods exist that constrain natural mortality estimates for all fish species in the world, the steepness of stock recruitment curves is not established for more than a few well-studied species. Thus, avoiding the use of this parameter is a plus.

R1.6: The CPUE data as presented in the appendix are extraordinarily clean. Normal data is much noisier.

R1.6 Answer: Reviewer 1 presumably refers to the simulated data. The CV used there for the random error for surplus production was 0.2; the CV for the random error for catches was 0.1. Given that many "full stock assessments" treat catch as free of error, we don't think that the simulated data were "extraordinarily clean". In any case, the simulations were only used to verify that the AMSY model works in principle; we made a point of evaluating AMSY against 140 real stocks with a wide range of life histories and environments, much more then are usually tested when a new method is presented.

R1.7: Line 155: another unjustified constraint that will improve the apparent performance of this method.

R1.7 Answer: Reviewer 1 refers here to built-in rules for the minimum and maximum range of relative carrying capacity k*q. These specify that the lower bound of carrying capacity must be equal to or larger than the maximum observed CPUE, and that the upper bound must be larger than the lower by at least 30% and max 300%. For example, in a stock where the max CPUE was 1000 tonnes, the prior for the upper bound of carrying capacity would be not less than 1300 tonnes and not more than 3000 tones. This is not a small range, and the tests against simulated stocks with known ranges showed this assumption to be realistic.

R1.8: Line 165-170: Justification for including this reduced productivity at low stock size is missing and it just adds to the sense that this model is a fantasy. Then, in the discussion, the authors rationalize not using other forms of the biomass dynamics approach which might be more justified then their modification.

R1.8 Answer: This comment refers to the built-in reduction of expected recruitment at low stock sizes. That recruitment is likely to be impaired at very small stock size is well known and widely used in management. For example, ICES reduces the applicable value for FMSY linearly when B < MSY Btrigger, which is very similar to our approach. We have added that reference.

R1.9: Line 172: productivity process error should have been identified in the earlier equations.

R1.9 Answer: We have added a statement that error terms were not shown in the previous equations for the sake of simplicity.

R1.10: Line 183-242: these filters to exclude unrealistic results create even more concern in my mind regarding the value of this approach.

R1.10 Answer: Monte Carlo methods are a well-established statistical procedure that uses the process of repeated random sampling to make numerical estimations of unknown parameters. Excluding parameter values that result in unlikely predictions, such as negative catches, is a logical addition to random sampling and is at the core of the new data-poor stock assessment methods such as CMSY. We have added a new section that explicitly looks at the influence of the filters on the results and we have reduced the number of filters from 7 to 5.

R1.11: Line 263: "This first year biomass was also used as prior for AMSY" seems like another trick to get the method to perform well.

R1.11 Answer: This refers to the time series of 50 years of simulated data, where the biomass prior was set to the 'true' biomass range (0.15 – 0.4 or 0.5 – 0.85) at the beginning of the time series. These are pretty wide ranges (63% - 41% of the max value). The ranges had to include the 'true' value for this test (otherwise they would not reflect independent correct knowledge about the stock) and they were taken 50 years before the value that would be compared with the AMSY estimate, so this is certainly no trick to make the method perform well. The text did not mention the ranges, which could have misled the reviewer. This has now been changed.

R1.12: Line 275: "…AMSY used the meanwhile available…" does not make sense.

R1.12 Answer: The reviewer has a point that the text was not clear enough. It was meant to say that during the previous BSM exercise FishBase only provided qualitative information about resilience. As part of the response to the reviewers, we have now re-run the BSM estimates and both BSM and AMSY use exactly the same prior distributions for r, based on prior knowledge available in FishBase (www.fishbase.org) or SeaLifeBase

(www.sealifebase.org). The point is to explore whether AMSY can get close to the results of BSM, using exactly the same priors for relative biomass in the first year and for r, but with the difference that BSM uses time series of catch and CPUE whereas AMSY uses only CPUE. This has now been made explicit in the chapter about how AMSY works and is also stressed in the Conclusions.

R1.13: Line 305-309: This appearance of good performance against simulated data is not surprising given the large number of constraints and priors used in the AMSY method.

R1.12 Answer: Two priors and 5 filters applying general population dynamics logic actually are not that many inputs compared with other data poor (see e.g. DCAC of MacCall above) or data rich (e.g. M and steepness) models. Note also that the resilience prior is derived 'objectively' from online databases and that the 5 filters operate on general rules that are built into the code, so potentially subjective user input is only required for the biomass prior. This is less than the usual priors needed for stock assessments. However, Reviewer 1 has a point in that the filter settings were fine-tuned with test runs against the simulated data; so, a comparison of results is not fair and it was removed from the text. The simulations provide only a proof of concept whereas the testing is done against 140 real world stocks. This is now made clear in the text.

R1.13: Line 311-312: If the BSM in Froese et al. (2017) uses any of the constraints and priors used for AMSY, then the similarity of their results is not surprising and might be considered simply an artifact of the methods

R1.13 Answer: The full Bayesian Schaefer model in BSM uses the same prior for resilience, but none of the filters. Instead, catch is a required additional input. The point here was to see whether AMSY can reproduce the BSM results WITHOUT knowing the catch. That is the case in the simulations even without the filters (this demonstration is new) and with most of the 140 tested real stocks.

R1.14: Table 3, Line 356: The confidence intervals are rather tight in comparison to many "full stock assessments" and there is an obvious correlation between the estimated F/Fmsy and B/Bmsy. Both of these observations are another indication that the method is too stiff with its many priors and constraints.

R1.14 Answer: Confidence limits of F/FMSY are actually rather wide, which is the reason why F/FMSY results of AMSY were explicitly NOT recommended for management in the text, in the Conclusions, and in the Abstract. In contrast, the AMSY estimates of B/BMSY have confidence limits that are similar to regular stock assessments. This point has been made more explicit in the text. Also, we have picked up on the correlation between r and k, which is also stressed by Reviewer 2, and have replaced the uniform prior distribution with a multivariate one and rerun all analyses.

R1.15: I fear that use of this method will give a misleading sense of knowledge that will not necessarily result in better management.

R1.15 Answer: We agree with Reviewer 1 about the danger that insufficient or wrong knowledge may be regarded as good only because it has run through a 'fancy' model. But this is a danger that applies to all models and is not specific or more pronounced in AMSY. We have tried to address this problem with a long section about properties and assumptions of AMSY.

Reviewer 2 Comments:

R2.1: Overview: The authors present a new method for applying limited fisheries data to provide science-based fisheries management guidance. Specifically, the AMSY approach uses catch-per-unit effort data as a proxy for catches and applies it in a Schaefer production model in order to estimate relative stock status in

terms of biomass and fishing mortality. This method occupies a unique spot in the spectrum of data-limited methods by using CPUE to access surplus-production based reference points in order to interpret stock status. This method frees the user from having to have a catch time series, which is a common issue in many fisheries. Overall, I think the idea of this method is a very valuable one for adding to the data-limited toolkit. Regarding the presentation of this method in this paper, I found many places where more work should be done, and more considerations should be made. There are many assumptions associated with this approach, only some of which are discussed, and none that are performance tested. Below, I outline several areas I believe the paper should look to improve before publication. I hope the authors find these suggestions useful and constructive, as I think the method is a worthy contribution to the growing data-limited methods literature.

R2.1 Answer: We are pleased that Reviewer 2 shares our vision and will strive to do as many of the suggested improvements as possible in the context of this 'proof of concept' paper.

Major considerations:

R2.2: While the concept of AMSY is interesting, the performance testing is really only proof of concept. There are two main performance testing approaches and then application to real data. The first approach is to simulate data and parameters, apply that information to AMSY, then compare to the known simulated values. While this is a typical approach to simulation testing, all that is tested here is whether you get the right values back if the right values are given to AMSY. What is completely missing and definitely needed is robustness/ sensitivity testing. There are many assumptions being made by this model (all input values being correct, CPUE not being biased or highly imprecise (lines 429-430; 445-447), q is constant, Schaefer model is appropriate for all life histories, etc.). Violating these assumptions and by how much to degrade performance are the true test of how this method performs.

R2.2 Answer: Yes, the simulations are a proof of concept but the simulation testing does more than just getting "the right values back if the right values are given to AMSY." The simulations tested extreme situations that are rarely encountered in real world situations, such as exploring results for species with very low resilience or stocks with very light exploitation. In both cases AMSY results were found less reliable, resulting in a warning to users. This warning was checked again and made more explicit.

R2.3: Likewise, the second approach is comparing the results to a surplus production approach that uses CPUE and a catch time series. I assume all the input parameter values for each comparison are the same between both models, thus it is unsurprising they perform similarly. The differences come down to filtering treatments in AMSY and maybe slightly different priors. I did not find this performance testing convincing of the utility of AMSY beyond that it can mimic other surplus production models WHEN given the same input values and relative catch data that have a direct relationship to abundance and catch. A proper simulation test exploring these issues, or a much-expanded Discussion section talking about all of these assumptions and offering caution in applying this method until proper simulation testing is done is needed.

R2.3 Answer: Apparently we have not presented the case clearly enough, since both reviewers take issue here: The test was exactly whether AMSY, which is a "half" Schaefer model without any information about extractions (=catches) can approximately reproduce the parameter estimates of a 'full' Schaefer model with catch as input, everything else being equal. This statement is now made explicitly in the text. What more can one expect from a data-poor method? Reviewer 2 has a point that catch and CPUE are strongly linked in the simulations, with deviations coming only from the error terms, but this is not the case for the 140 real stocks, where variations in catch can stem from management or economics or other drivers and similarly, CPUE can be driven strongly by environment and less by catches. Still, AMSY was able to approximate most of the BSM predications. We have revisited the respective text and made it more explicit, in M&M as well as in the Discussion.

R2.4: In the description of the filters, there are several places (lines 191, 200, 211, 215, and 222) where values are "derived from simulations". What are these simulations? It is not possible to tell how much uncertainty there is around these values, nor if these values are universally applicable. There needs to be further description and reporting on what these simulations are in order for the reader to understand the derivation of these filters.

R2.4 Answer: We have now made more explicit how the parameterization of the logical filters was done (basically, test runs against the simulated data) and that they are preliminary. However, they worked reasonably well on the wide range of 140 real world stocks and thus seem fit for preliminary use. This is now stated explicitly in the text.

R2.5: In evaluating the need for so many filters and the fact that the k_q prior is derived from Bt/K, it does not seem like the correlation structure for r-k is considered. If it was, I am curious if this would decrease the number of needed filters. And given the model structure, that correlation structure is needed. Please check to see (and possibly report) the posterior relationship between r and k_q to make sure the needed correlation structure is present. If not, try to add correlation structure and see if that decreases the number of triggered filters, and therefore the possible reduction in needed filters.

R2.5 Answer: Following this recommendation of all reviewers, we have now implemented a multivariate log-normal distribution based on the posterior correlation of r and k in 140 real stocks. This is described in a new section in the text. All analyses were subsequently redone, with no major changes in results but slight improvements and the opportunity to remove two of the previous 7 filters.

R2.6: Lines 244-252: The approach of trimming the r-k_q pairs further via the MSY_q distribution, then adding another 30% uncertainty (is this CV?), is strange and seemingly arbitrary (why 95% CI?). Why is this done? Why not just stick with the output distributions? And how do you know what "unrealistically narrow" CIs are?

R2.6 Answer: We have removed this trimming and now use median and quantiles of the results for most likely central value and approximate 95% confidence limits.

R2.7: Line 273: What biomass (total, spawning, other?) was used from the stock assessments? This matters as to the interpretation of what the surplus-production model is measuring, and how the underlying selectivity assumption matches the assumption of FMSY=r/2. This is needed to understand Table 2 results as well. Please clarify the biomass and acknowledge the selectivity assumption.

R2.7 Answer: About half of the examined stocks reported CPUE data from commercial fishers or from scientific surveys and the other half reported spawning stock biomass, which was treated as CPUE. For the purpose of this study, the type of CPUE was irrelevant because both BSM and AMSY used the same CPUE data, and the purpose of the exercise was to see whether AMSY can reproduce the results of a full Schaefer model without the catch being known. In other words, problems resulting from TSB versus SSB versus commercial CPUE versus survey CPUE would have affected both BSM and AMSY.

R2.8: On line 418, it is mentioned that selectivity is contained in the r and k parameters, which is not correct. In this case, it is contained in the CPUE.

R2.8 Answer: We agree here with Reviewer 2 and have removed the mention of selectivity in that sentence.

R2.9: Lines 361-375: The argument for using the Schaefer model is not strong. There is a good understanding, based on life history, what BMSY might be. The Pella-Tomlinson is a general formulation, thus giving a much more flexible approach, whereas using the Schaefer model may significantly overestimate overfishing. Why not allow the flexibility and have the user specify the shape parameter?

R2.9 Answer: Reviewer 2 prefers the flexibility of the Pella-Tomlinson model. However, there is no biological basis to its additional shape parameter, i.e., what values it should take as prior and thus how the deviation from the logistic S-shaped curve of population growth should be, based on the life history of the species. Given that we are dealing with a data-poor situation, it does not seem prudent to try to estimate this ill-defined parameter. We have stressed this point within the existing justification for selecting the Schaefer model.

Minor considerations:

R2.10: There are several places where values are called reference points when they are not. Line 46: F/FMSY and B/BMSY are not reference points. FMSY and BMSY are reference points.

R2.10 Answer: We agree and have fixed the text accordingly. See also R2.19.

R2.11: Line 80: Schaefer model assumes references points (e.g., FMSY=r/2), it does not calculate them. Again, I think stock status is what is meant.

R2.11 Answer: While in our opinion the Schaefer model does estimate r and k, we followed the reviewer and replaced “reference points” with “stock status and exploitation”.

R2.12: Line 295: r, k_q and F are not reference points.

R2.12 Answer: F was not in the list; r, kq and MSYq are now referred to as “population dynamic parameters”.

R2.13: Line 168: The value of 0.25 seems arbitrary. What is it based on? Should this be stock specific?

R2.13 Answer: Half of BMSY is widely accepted as a proxy for Bpa or the border of safe biological limits below which recruitment may be impaired. In the Schaefer model 0.5 BMSY = 0.25 k. The text has been amended to make this connection clear, with reference to the ICES advice background, i.e. their rule for reduction of F when B < MSY Btrigger.

R2.14: Lines 240-241: Does this also happen in CMSY when fitting catches? It seems this may be a sign of parameter misspecification. Double check that this isn’t a sign of something to address.

R2.14 Answer: This is a misunderstanding. AMSY assigns a lognormal random observation error to the input CPUE data. Thus, different runs have different observation errors for every single year in the time series, and only those parameter values and ‘observation error-corrected’ CPUE trajectories that pass all filters contribute to the results. The text has been amended to better reflect this process.

R2.15: Lines 315-316: BSM had higher k and r variance, but this could be due to the CV on the index being fix by the BSM model (same as AMSY?) and the prior on k, which is not clear if it is the same magnitude as the prior on k_q. And the fact that AMSY is more precise is not necessarily a good thing given that Bayesian models are usually better at estimating uncertainty.

R2.15 Answer: Lines 315-316 do not refer to BSM variance, so it is not clear what Reviewer 2 is referring to here. In any case, as stated, AMSY variance of F/FMSY is higher than in BSM and B/BMSY variance is about the same or also higher, with few exceptions. Since this may be a misunderstanding and since this is listed among minor issues, no action was taken.

R2.16: Lines 326-333: It is not obvious if comparing a relative to absolute measure (k_q vs K and MSY_q vs MSY) is a fair comparison as the absolute value is also taking into consideration the value of q. Please consider if this is appropriate.

R2.16 Answer: Reviewer 2 has a point, because the observed deviations may in part be caused by catchability q, which is not taken into consideration by AMSY. We have added a sentence that makes this potential source of divergence explicit.

R2.17: Lines 346: One of the most interesting results is that the independent LBB [Froese et al. 2018] approach and this approach may be getting similar stock status estimates. While this is not always expected as it is very common for data types to contradict each other, to be able to confirm this behavior would be good, but I do not see this comparison in the paper. Are readers expected to go and read the two references in order to make this comparison? Why not add it to Table 3?

R2.17 Answer: We agree with Reviewer 2 that the good agreement between trends in B/BMSY estimates based on length frequencies and on CPUE are unexpected and very encouraging. We do not stress this more because LBB results (for a selected year) were used as priors for AMSY, so the AMSY results are not fully independent of LBB. Comparing the similarity between LBB and CPUE trends is beyond the scope of this AMSY paper, but in a way we raise a flag here for subsequent papers to explore this aspect. No action was taken.

R2.18: Lines 407-408: I do not agree one cannot use uncertainty estimates of fishing rates. By this argument there emerges some arbitrary line of what is too much uncertainty. The authors' sentiment implies that uncertainty should not be ignored, which is an important point. One therefore needs to express their risk tolerance in order to handle the uncertainty. I would remove this rejection of use and make it a strong consideration/warning on how to handle the uncertainty in this metric.

R2.18 Answer: We agree with the reviewer and have changed the text accordingly.

Edits/suggestions:

R2.19: Line 47: Biological reference points, thus those associated with MSY, are based on life history values (r, M, Linf, k, maturity, etc.) and selectivity, thus they often can be estimated with limited data. It is stock status (values relative to the reference points) that is typically unknown. This sentence should probably just refer to stock status and not reference points to maintain its main message.

R2.19 Answer: The text was changed to: "…exploitation level and stock status are unknown…". This also solves R2.10.

R2.20: Line 58: Claiming that CPUE and abundance are in good agreement is loaded with assumptions, and thus should not be stated as matter of fact. Some of these issues are addressed in the Discussion section; having such a broad statement here is unnecessary. Just saying that CPUE can reflect abundance under certain conditions should be good enough.

R2.20 Answer: The text has been changed to reflect this concern.

R2.21: Line 72: "the resulting presentation" is an odd turn of phrase, and I am not totally sure what it means. Please consider revising this to be clearer.

R2.21 Answer: The text was rephrased to make this point clearer.

R2.22: Line 90: "… resulting in a factor of 1 when Bt=0…". Done.

R2.23: Line 91: " CPUE is often assumed…". Done.

R2.24: Line 126: A prior for k_q should also be mentioned here, even though it is derived later from Bt/k.

R2.24 Answer: As the reviewer says, a prior for kq is derived without intervention by the user by combining the relative biomass Bt/k prior with the observed CPUE in that year. Thus, it would not be correct to state that a prior for kq is required input in addition to a prior for relative biomass. Given that the derivation of the kq prior is explained in detail in the text elsewhere and was understood correctly by reviewer 1, we left the text here as is. See also R3.5 Answer.

R2.25: Line 134: Need more details on how r was derived from FishBase. What section? Life history tool or somewhere else?

R2.25 Answer: Done.

R2.26: Figure 1 font is very small. Just want to make sure this improves for publication.

R2.26 Answer: Figure 1 will be redone with larger fonts prior to publication.

R2.27: Line 389: "in the penultimate year were…".

R2.27 Answer: Done.

Reviewer: 3

General comments

The authors present a novel approach for the assessment of data-limited stocks. Their method is applicable to stocks for which abundance trend information (CPUE) is available, along with estimates of stock productivity and at least one independent estimate of relative stock status (B/K). Catch data are not required. The discrete Schaefer model for biomass is recast into units of CPUE and 'relative catch' (C[q,t]) with associated relative reference points (K[q], MSY[q]). The manuscript is easy to follow, but in general does not adequately describe how the results are sensitive to prior specifications and the chosen set of filters based on the CPUE time series.

Here are a few topics I'd like to see addressed in a revised manuscript:
R3.1: The method requires a prior for relative biomass (B[t]/K); however, the text does not clearly define how to choose a year for this prior. The LBB approach requires length-frequency data, so it seems logical that the priors developed in this manner would apply to a year during which the lengths were collected. Was that the case? Other parts of the text refer to the prior being applied to the first year in the time series (e.g. the "Simulated Data" and "Real Data" sections). When and why was the first year chosen for the prior?

R3.1 Answer: Text was added in the section 'Priors for r, kq and F/FMSY' to explain the selection of the best year for the Bt/k prior.

R3.2: The results should better describe how the proposed filters affect the results. Currently, the authors state that the upper or lower end of r & k are affected, but it would be useful to include marginal density plots for key parameters as well as for derived quantities (F/FMSY and B/BMSY) before and after the filters were applied.

R3.2 Answer: We have seriously considered this proposal; however, the effect of the filters can be very different for different simulated or real stocks, basically being a function of r and CPUE/kq. Exploring and showing comparative density graphs for all possible combinations of r and CPUE/kq is an extensive paper in its own right, which some of the co-authors are keen to do, but it is clearly beyond the scope of this 'proof of concept' presentation of a new method and the interest of most readers. We have, however, added a section where we explain better the purpose and function of the filters and suggest that this is an area where AMSY will benefit from future research. Note also that the current paper already far exceeds the testing done in other widely accepted presentations of new data-poor methods. For example, MacCall's DCAC method mentioned as good example by two of the reviewers uses two examples of fully assessed stocks to demonstrate the usefulness of the method. In comparison, this study uses 24 simulated stocks for proof of concept and 140 fully assessed stocks for evaluation.

R3.3: It's unclear whether the relative reference points (F/Fmsy and B/Bmsy) are sensitive to changes to the assumed values of process and/or observation error. A set of values were chosen (lines 171-172), but there is no evaluation of how F/Fmsy and B/Bmsy are affected by these assumptions. One option is to draw from distributions of the variance parameters, examine which values are not excluded by the filters, and plot the retained values against model outputs.

R3.3 Answer: We have stressed that the chosen values are preliminary. They can be changed by the users.

R3.4: The supplementary materials (>200 pages!) as currently presented will be overwhelming to most readers without further clarification. Appendix 3, in particular, would benefit from a 'walk-through' of the method for a single stock. The 1/2 page of text at the beginning of Appendix 3 is inadequate to allow for replication of the study.

R3.4 Answer: We note that the reviewers ask for substantially more testing and details and then complain about the presentation of the testing that was already done being too long. But we followed the thrust of this comment and have revised the text in the Supplement.

R3.5: The prior for K[q] is derived from the prior for B[t]/K (line 150) and an observed value of CPUE (at time t). How is t chosen, and doesn't that suggest that the prior for K[q] is simply a transformation of the prior for B[t]/K? Bivariate plots showing the correlation structure of the joint prior (r, B[t]/K, and K[q]) would be useful to understand how these priors are correlated.

R3.5 Answer: The choice of the year of the prior is explained now in the text as answer to R3.1. If one has an absolute stock size Bt/k and a relative stock size CPUEt for the same year, then one can calculate the range that kq can take from kq = CPUEt/(Bt/k), by inserting CPUEt and using first the lower and then the upper range value of the Bt/k prior. This is already stated in the text below Table 2, and in so far the prior for kq is directly derived from the Bt/k prior. In other words, the prior for Bt/k is used to preliminary put the CPUE into an MSY framework. This placement is then refined by the AMSY Monte Carlo filtering process. A text to this effect has been inserted in the paragraph below Table 2. See also R2.24 Answer.

R3.6: Details of the filtration process are not adequately described (see below), and the authors should include better descriptions of these filters in the text and not rely on links to the code. The fraction of MC simulations excluded by each filter should be included in the results.

R3.6 Answer: This concern has been answered under R3.2. To address it, we have added new text to better explain the purpose and effects of the filters. Also, we have removed two of the filters.

Line-specific comments

R3.7: Lines 155-156: In this case, it suggests that the prior for B[t]/K is poorly specified.

R3.7 Answer: This case refers to e.g. a kq prior close to unexploited, e.g. 0.8 – 1.0. The rule for minimum kq prior range will expand this to 0.8 – 1.04. In an opposite scenario, consider a CPUEt = 100 and a Bt/k prior of 0.01 – 0.4, giving an upper kq prior range of 100/0.01 = 10000 and a lower range of 100/0.4 = 250. The rule of maximum kq prior range will reduce the upper range from an unrealistic 10000 to a more realistic 3*250 = 750. Yes, one could say that the lower Bt/k prior range of 0.01 is poorly defined by the user, but it is a common way to reflect uncertainty about stock size, and the built-in rules can deal with that. Since the two other reviewers had no problems here, no action was taken.

R3.8: Lines 181-190: Is 'catch' actually 'relative catch' in this paragraph? The degree of negative catch allowed (7% of K) is said to be based on simulations, but that is probably sensitive to the assumed values of process and observation error variance. A better description of the rationale used to identify the 7% is needed for users who may select different values for the error terms.

R3.8 Answer: In the context of AMSY, these are indeed relative catches and that has been corrected in the text (note that the logic of the argument presented there does not change). We revisited this number and tried different values for allowed negative relative catch in test runs against the simulated data. New values are -6% for very low and -2% for low productivity. The main purpose of the study is to present a new method with preliminary tests, not to explore optimization of all parameters. That should be the work of subsequent publications, as has been done with other assessment methods. To accommodate this criticism, we stress the need of additional sensitivity testing in the Discussion. See also R2.24 Answer and R3.5 Answer.

R3.9: Line 197: It's not clear what is meant by "empirical multipliers (based on simulations)". Each filter should be described in detail, as this is the basis for how the model excludes parameter combinations from which the reference points are derived. Line 207: same comment -- "MSY-multiplier values derived from simulations"? Line 225: The boundaries (-3 and 5) used to identify 'unrealistic' values of F/Fmsy are going to have a greater or lesser effect depending on the user's choice of process and observation error terms. How were these values chosen?

R3.9 Answer: These points are addressed in R2.24 Answer, R3.5 Answer and R3.8 Answer.

R3.10: Lines 247-248: How is this enforced? Are additional MC simulations generated from an assumed error distribution with the same mean?

R3.10 Answer: Reviewer 2 also questioned whether this additional filter was needed, and we removed it. See also R2.6 Answer.

R3.11: Lines 258-259: "...first year biomass was also used as a prior..." [see comment R31 above]

R3.11 Answer: The text was amended to explain that the first year was chosen to minimize influence of this prior on the estimated relative biomass 50 years later.

R3.12: Lines 288-303: How were the AMSY priors specified relative to 'true' values? Was AMSY given priors with the same central values at the 'true' values?

R3.12 Answer: No central or 'true' values were used as part of the priors, only uniform ranges which included the true value somewhere.

R3.13: Lines 305-307: How were the relative biomasses from AMSY converted back to unexploited biomass estimates from the BSM model? This would require knowledge of the true q value, wouldn't it?

R3.13 Answer: In the simulations, the 'true' values for kq and MSYq were known (no transformation needed) and could be used for the comparison of results. Reviewer 3 has correctly noted that instead the text wrongly referred to k and MSY. We have corrected that error and thank reviewer 3 for having spotted it.

R3.14: Line 338:

R3.14: Answer: Line 338: see general comment #4. See R3.4 Answer.

R3.15: *Lines 354-367: As Maunder (2003) pointed out, it is preferable to use the Pella-Tomlinson model and specify values for the shape parameter (i.e., making the assumption regarding Bmsy/k explicit). Values of Bmsy/k greater than 0.5 have also been proposed (Walters and Kitchell, 2001; MacCall 2002), and the P-T model would allow the user to explore these alternatives and their impacts on results.*

R3.15 Answer: Maunder (2003) states: "This shape parameter can in theory be estimated from time series data, although in practice, there is generally insufficient information in individual data sets to estimate it with any precision." [..] "Such studies combine life history theory with assumptions about individual growth parameters, fishery selectivity, and other relevant age-structured effects to provide bounds on plausible values for SB MSY reference points. However, the use of such strategies to inform the shape of the surplus production relationship remain highly contested [..]." Given that AMSY is a data-poor method, and given that there are no widely accepted and easily accessible priors for the shape parameter, we followed general scientific practice and used the most parsimonious available model, i.e., the logistic curve of population growth and the resulting Schaefer model for surplus production. We have extended the text of the discussion of this topic.

R3.16: Lines 375-377: This is difficult to interpret without a better description of how AMSY was configured in the simulation study (see comment for lines 288-303) See R3.12 Answer.

R3.17: Line 461: Fmsy = r/2 because of the assumed Schaefer model; variability in Fmsy is underestimated because the model assumes Bmsy/k = 0.5.

R3.17 Answer: See R3.15 for the reasoning for selection of Schaefer over Pella-Tomlinson. Note also that uncertainty in F/FMSY estimates of AMSY is very high and given as a reason using these estimates with caution (if at all). The main result of AMSY, as stressed in several places in the text, is an estimate of recent stock status (B/k).

Discussion

While responding to the many points of the reviewers was tedious and changing the model and re-running about 200 analyses was a lot of work, there is no doubt among the authors that the new method was substantially improved by the review process. The authors use this opportunity to thank the Editor and the 3 reviewers for the considerable work they have invested in that process.

References

MacCall, A. 2002. Fishery-management and stock-rebuilding prospects under conditions of low-frequency environmental variability and species interactions. *Bulletin of Marine Science* 70(2): 613–628.

Dittert, N., M. Diepenbroek and H. Grobe, 2001. Scientific data must be made available to all. *Nature*, 14:393.

Froese, R. and R. Reyes, Jr. 2003. Use them or lose them: The need to make collection databases publicly available, p. 585-591 *In*: A. Legakis, S. Sfenthourakis, R. Polymeri and M. Thessalou-Legaki (eds.). *Proceedings of the 18th International Congress of Zoology*.

Maunder, M. 2003. Is it time to discard the Schaefer model from the stock assessment scientist's toolbox? *Fisheries Research* 61:145-149.

Pauly, D. 2018. Learning from peer-reviews: The Gill-Oxygen Limitation Theory and the Growth of Fishes and Aquatic Invertebrates, p. 53-70. *In*: D. Pauly and V. Ruiz-Leotaud (eds.) *Marine and Freshwater Miscellanea*. Fisheries Centre Research Reports 26(2). Institute for the Oceans and Fisheries, University of British Columbia.

Froese, R. D. Lloris and S. Opitz. 2004. The need to make scientific data publicly available – concerns and possible solutions, p. 267-271 *In*: E. Feoli and C.E. Nauen (eds.) Proceedings of the INCO-DEV International Workshop on Information Systems for Policy and Technical Support in Fisheries and Aquaculture. ACP-EU Fisheries Research Report No 8.

Froese, R., H. Winker, G. Coro, N. Demirel, A.C. Tsikliras, D. Dimarchopoulou, G. Scarcella, W.N. Probst, M. Dureuil and D. Pauly. 2018. A new approach for estimating stock status from length-frequency data. *ICES Journal of Marine Science* doi: 10.1093/icesjms/fsy078. [With corrigendum; same authors, title and journal: 2019. Corrigendum, 76(1): 350-351, doi: 10.1093/ices/fsy139]

Walters, C. and J. Kitchell, 2001. Cultivation/depensation effects on juvenile survival and recruitment: implications for the theory of fishing. *Canadian Journal of Fisheries and Aquatic Sciences* 58: 39–50.

Zeller, D., R. Froese and D. Pauly. 2005. On losing and recovering fisheries and marine science data. *Marine Policy* 29: 69-73. doi 10.1016/j.marpol.2004.02.003.