AE comments: We have received two reports from expert Referees. While both appreciated the timeliness and the topic of the manuscript, they also raised substantial concerns. I am issuing a "Major Revision" decision in order to give the Authors a chance to respond to the criticism. Some of the points raised by the Referees should however be fully addressed for the study to be published in PLoS Computational Biology. The appropriate literature should be cited and discussed, the novelty and generality of the approach should be made clear, and the method should be rigorously tested. In case of a resubmission, I will contact the same Referees, and potentially add another.

>Thank you for the opportunity to submit a revision of our manuscript. We would like to thank both reviewers and the AE for their thorough and insightful comments. We carefully considered all of the individual comments and addressed each of them. We have now added substantial text, including 15 new citations, to help contextualize the novelty and generality of the approach. We also re-read the entire manuscript and revised for clarity where appropriate. Please see our point-by-point responses to each of the reviewer’s concerns below.

Reviewer #1: In this paper the authors present a method to detect a shift in governing parameters in a time series. I have not seen such method for finding a shift in parameters and this method could be interesting for biologists. Not being a statistician I cannot fully judge if the method of selecting the optimal parameter shift is novel enough for PLOS Comp Biol.

>Thank you for your thoughtful comments and positive assessment of our work! We agree that our approach will be new to biologists, as it is a significant departure from previously proposed methodologies. The novelty of our approach is that it can offer new insights into time series data as it specifically addresses several issues associated with previous regime detection approaches, and thus has the potential for broad applicability. Unlike previous approaches to change-point analysis, the algorithm we describe does not just identify break points, but provides a means of weighting them to evaluate confidence in whether a break has occurred.  
  
My main concern with the paper is that the authors suggest that the method is suitable for testing shifts to alternative stable states, while they do not test that. Only in the abstract and introduction this is suggested, after that they call it a test for parameter shifts (I agree with that term). Although the term “regime shift” is not so well defined, it often refers to a shift between alternative attractors. (see for instance Anderson et al. 2009 (Trends in Ecology & Evolution, 24:1 49-57)). When a system shifts between alternative attractors there is not a clear shift in external conditions needed, but just a small external change may make the system cross a tipping point. That is clearly not what this method is testing for as the authors test for a clear shift in the parameters of a model. From the title I expected that this paper was about detecting such tipping points. To avoid such confusion I think the title can better be changed and the authors should explicitly define “regime shift” and state that the method is not for detecting tipping points. There is an extensive literature on detecting tipping points (see some examples below) and I think the authors should discuss the difference between those methods and their parameter shift detection method.

>The reviewer brings up a very good point. There is sufficient ambiguity around terminology reflecting abrupt ecological transitions that Bahlai organized a symposium and panel discussion to try and synthesize these concepts at last year’s Ecological Society of America meeting (<https://eco.confex.com/eco/2018/meetingapp.cgi/Session/14351>). But clearly, from this and Reviewer #2’s comments, we had not sufficiently covered this subject in the introduction of the paper. We restructured the introduction to focus on this issue in a new second paragraph, and focus on the specifics of density dependent population modelling later, to provide a more inductive framing of our paper’s goals.

We modified our terminology throughout the manuscript to clarify that our algorithm is designed to detect shifts in parameter estimates (and not dynamic states). We also updated the algorithm’s name to the “Dynamic Shift Dectector” and changed the title of the paper for increased clarity.  
  
For detecting tipping points there are many more methods and literature that was not cited in this paper. Only the 2000 paper of Hare and Mantua was cited and not the papers by Rodionov (among others “A sequential algorithm for testing climate regime shifts”) and the review of Anderson et al. 2009 (see above) and various marine regime shifts (among others: Weijerman et al., 2005. Regime shifts in marine ecosystems of the North Sea and Wadden Sea. Mar. Ecol. Prog. Ser. 298, 21–39; Beaugrand 2004. The North Sea regime shift: Evidence, causes, mechanisms and consequences, Progress in Oceanography, 60, Rocha et al. 2015 Marine regime shifts: drivers and impacts on ecosystems services).

>Thank you for bringing these papers to our attention. We included several of them in the introduction, and included a paragraph tying our work to the Andersen paper specifically (see introduction). The Andersen paper was especially useful as it notes that change-point analysis approaches suffer from limited sensitivity and precision due to their frequentist statistical approaches, and thus could be improved using a model selection approach (as we do in our paper). This insight was particularly useful in helping us connect our current work with previous work on the topic.

I don’t get why the authors only use the Ricker model. I don’t understand the argument that “these methods [breakpoint analysis] do not work on data with internal, density dependent structure“. Any system can have transient dynamics, what is so special about populations? Why is the Ricker model so generic, while other methods are ad hoc? In populations there is not a single “true” model. The Ricker model for example does not account for any interactions between populations.

>This is a good point. It is not our intention to make this paper ‘about’ the Ricker model. The Ricker model is simply a model that we know fits a variety of population time series data and structures (including that from our case studies). The Ricker model provides a more complex and ecologically meaningful approach than using more simplified moving average/variability measures. However, we argue that Ricker is a good starting point because its shape is flexible, allowing for near linear growth, compensatory, and overcompensatory dynamics. Although these are not the only possible options for population growth, they are arguably the most commonly observed in population data. We have added an explanation to the introduction to address that our proposed algorithm is itself agnostic to a specific model-structure and that we simply use the Ricker model as a starting point and to test the method.

On the other hand, the method can easily be generalized to any population model. Especially in situations where there are clear interactions between populations, I think it would be better to model them too. I would be nice if the R package could be used for other models too.

>We agree that our framework is compatible with a broad range of population models and could even be extended to interacting populations (if such data were available). The next steps in the development of this method are to extend the approach and turn it into an R package, which would allow for user-inputted model structures. While we provide all of the R code used in this paper on our Github site with detailed tutorials, creation of an R package is beyond the scope of the current paper.  
  
A known issue with the Akaike Information Criterion is that some authors think that it is too liberal in accepting extra parameters and that there is a risk of overfitting. The authors acknowledge this risk and use the Akaike weights (a measure of the relative improvement compared to the optimal model) to deal with that. As these are not so well-known for biologists, I think this should be explained in the methods (also the AICc).

>Another good point. We have added additional text to the methods explaining Akaike weights. Although this section is short (because we do not want to distract from the core model methodology), we think this will be useful to readers unfamiliar with weighting approaches. We also added a paragraph in the discussion to address the implications of metric choices on the performance of our algorithm.  
  
Reviewer #2

My recommendation

*Major revision. The study fails to fully address how the findings relate to previous research in this area. The authors should rewrite specifically their Introduction and Discussion to reference the related literature. Moreover, the authors should clarify the aspects criticised above to avoid confusion. And the technical details should be expanded and clarified to ensure that readers understand exactly what the researchers studied.*

>Thank you for your comments and thorough review of our paper. We agree with your suggested edits and have made major revisions to the paper accordingly. Specifically, we have rewritten a significant part of the introduction contextualizing the model and its place in the literature. We also added text to the discussion section and have worked to clarify the methods. Please see our responses to each of the specific comments below.

1. Summary of the research

*The article "The Regime Shift Detector: an algorithm to identify changes in dynamic rules governing populations" by Christie A. Bahlai1 and Elise F. Zipkin (PCOMPBIOL-D-19-00395) deals with the detection of regime shifts using a regime shift detector model to identify break points, which is based on the non-linear Ricker curve/function using Akaike’s Information Criterion corrected for small samples (AICc). Beside several other functions (such as the Beverton/Holt model or the segmented regression) the Ricker model originates from fisheries sciences and is used there to relate the advent of recruits to the previous year’s spawning stock biomass (SSB, lagged by one year), among others of cannibalistic fish populations with a negative stock density effect (such as cod or pikeperch). Hence in fishery sciences two different life stages of fish are related by these models: Recruits (juvenile fish, larvae, eggs) usually in numbers and SSB usually in terms of biomass. However, the validation of the detector model has been performed and illustrated here using two terrestrial examples plus numerical simulations. Based on the results the conclusion of the two authors is stating a good performance of their detector model based on a detection rate of around 70%.*

>It seems both reviewers leave with the impression that our paper is ‘about’ the Ricker model, which was not our intent. Although we use the Ricker model to illustrate how the algorithm can be used, the DSD algorithm is capable of accommodating many different model structures. Because the Ricker is a very flexible distribution, it can be used to accommodate a variety of functional forms, including within terrestrial systems. Please see our response to Reviewer 1 on this topic and the revised introduction that clarifies that point.

*In the light of global and climate-change the topic of detecting regime shifts is a rather important and quite popular one with many articles being published so far. Some of these publications deal with the definition of regime shifts alone by presenting lots of examples in various fields and areas (terrestrial and marine), others offer and provide shift detection methods (partly independent of the nature of regime shifts), while other publications are dealing with a combination of the two.*

>We agree with the reviewer about the importance of being able to detect changing regimes in light of the many stressors that animal populations face from a changing world. As suggested by both reviewers, we modified the introduction to further highlight the many publications that have considered this topic as well as the current gaps with such methods and the uniqueness of our approach.

*However, this complex topic is of high interest not only in biology but preferably in economics why a vast number of methods have been developed and derived in the field of econometrics dealing with interrupted time series, break point analysis, intervention models etc. to identify single or combined events (for instance, impulses based on oil crises, economic crises, financial crashes, earthquakes and other hazards, etc.) by incorporating them into a single time series model, sometime with multiple factors (transfer and intervention functions, vector autoregressive models, etc.). In fact, the evolutionary history of these type of methods show that most of them originate from economics and econometrics, respectively. Other areas where those methods have been developed are atmospheric physics and meteorology. Compared with these two fields (i.e. economics and meteorology) the use of high sophisticated methods in biology became popular rather late, not least because a professional training in statistics, informatics and numerical mathematics played only a little role in biology in the past (and even today).*

>Thank you for pointing out the vast literature in other fields that is relevant to our methodology. We have reviewed the broader literature and incorporated many papers from other fields on this topic (including a number of those suggested below).

*My overall impression of this article is that the authors do not address, illustrate and discuss the complexity as well as several aspects of their topic sufficiently and thoroughly enough. This holds for the definition of a regime shift in principle (in contrast to the definition of a “normal” shift or change) and the context of their method which has not been contrasted well enough with other alternative and competing methods, respectively.*

> This is a fair assessment, and this comment helped us to identify the points where our core message was not being realized. It seems both reviewers leave with the impression that our paper is ‘about’ the Ricker model, which was not our intent, and also that we did not provide enough historical context about shift detection methods. We added additional information about why the Ricker model was used (as an example of a relatively simple nonlinear mechanistic process that could be fit easily to accessible data, rather than more simple moving averages or the like to describe a system), and, as suggested by both reviewers, additional information to both the introduction and discussion contextualizing the dynamic shift analytical method. In response to both reviewers’ comments, we also modified our terminology throughout the manuscript to reflect concerns about the ambiguity in the use of the term “regime shift”, including modifying the algorithm’s name. Finally, we have highlighted the novelty of the study: change-point analyses suffer from limited sensitivity and precision due to their frequentist statistical approaches and thus could be improved using a model selection approach.

2. Examples and evidence

Major issues

*In the light of this, my impression of this paper is that the two authors are not aware of the many different methods and related vast literature and projects dealing with shift detection topics as they do not cite any fundamental articles from these scientific areas, including no econometric textbooks. Hence, the authors need to include / cite much more relevant literature and discuss their results in the light of these. Some examples that deal with shift detection in physics and biology that relate to econometrical or physical high-performance methods are for instance*

*1. Rodionov, S. (2004) A sequential algorithm for testing climate regime shifts, Geophys. Res. Lett., 31, L09204, doi:10.1029/2004GL019448.*

*2. Rodionov, S. N. (1994) Global and Regional Climate Interactions: The Caspian Sea Experience, Kluwer Academic Pub., Dordrecht, The Netherlands.*

*3. G. Beaugrand , A. Conversi , S. Chiba , M. Edwards , S. Fonda-Umani , C. Greene , N. Mantua , S. A. Otto , P. C. Reid , M. M. Stachura , L. Stemmann and H. Sugisaki (2015). Synchronous marine pelagic regime shifts in the Northern Hemisphere. https://doi.org/10.1098/rstb.2013.0272*

*4. R. P. Harris and J. H. Steele (2004) Regime shifts in the ocean. Reconciling observations and theory Volume 60, Issues 2–4, Pages 133-402*

*5. Gröger, J.P, Missong, M., Rountree, R. A. (2011) Analyses of interventions and structural breaks in marine and fisheries time series: Detection of shifts using iterative methods. Ecological Indicators 11 (2011) 1084–1092. doi:10.1016/j.ecolind.2010.12.008*

*6. Arula T, Gröger J, Ojaveer H, Simm M (2014) Shifts in the Spring Herring (Clupea harengus membras) Larvae and Related Environment in the Eastern Baltic Sea over the Past 50 Years. PLoS ONE 9(3): e91304. doi:10.1371/journal.pone.0091304*

*7. Beyraghdar Kashkooli O, Gröger J, Nuñez-Riboni I (2017) Qualitative assessment of climate-driven ecological shifts in the Caspian Sea. PLoS ONE 12(5): e0176892. https://doi.org/ 10.1371/journal.pone.0176892*

*8. Lindegren M, Dakos V, Gröger JP, Gardmark A, Kornilovs G, et al. (2012) Early Detection of Ecosystem Regime Shifts: A Multiple Method Evaluation for Management Application. PLoS ONE 7(7): e38410. doi:10.1371/journal.pone.0038410*

*9. Other authors and articles are related to the working group and publications of Jürgen Alheit.*

*10. The GLOBEC project:* [*http://www.globec.org/*](http://www.globec.org/)

>Thank you for this comment and for the work done to provide us with a list of relevant and useful literature. We reviewed the papers above and also spent time looking at the broader literature on regime shifts in fields outside of ecology and population dynamics. As a result, we considerably revised the introduction to include a more detailed literature review on this subject. We also incorporated many of the above suggested papers as well as several parallel papers from the terrestrial ecology literature.

*Apart from this, given a lot of existing competing methods, it is not clear to me*

• *What is the novelty of the study here?*

• *What is the added value/merit of this paper in the light of the many other sophisticated methods presented in the past so far? What is the difference to other existing methods?*

• *What are the advantages, what the dis-advantages when contrasting this work with that of other authors in this field?*

>Thank you for this comment, which helped us to explicitly clarify the utility of our proposed methodology. We now include more details on our reasoning and motivation for this work in the introduction. Briefly, we address a gap in how abrupt changes are assessed (identified by Bestelmeyer et al 2011 and Anderson et al 2009) in population processes. A lack of an adaptable, reasonably sensitive analytical method has resulted in many authors failing to directly address uncertainty around where break points assigned, with some applied altogether *ad hoc* in piecewise regressions. Here, we use a likelihood based, model selection approach to change point analysis, and we do so with a tool that can incorporate mechanistic models for ecological processes, in this case, a non-linear model for population regulation. The novelty of this method lies in the model selection approach used within the DSD algorithm, which allows for greater probability of detection than previous change-point type models, and the additional functions for weighting break points allows a user a measure of confidence in a given break point beyond 0/1. Previous methods of this type based in a frequentist statistical paradigm have limited ability to detect shifts in data of the size typical to contemporary long term organismal data (20-40 years) (Anderson et al 2009), and do not directly assess size of, or provide measures of confidence in-breaks.

*From this article it is even not clear how the authors define a regime shift in contrast to a significant shift or change in a time series? To my mind a definition of a regime shift has a philosophical and conceptual dimension and requires more features to explain than that of a “simple” (but significant) shift in a time series?*

>Thank you for this comment. In light of this comment and that of reviewer 1, we have now included some discussion of terminology around abrupt shifts. We define how we use the term ‘dynamic shift’ explicitly in the introduction. In this paper, we consider a dynamic shift as an abrupt change in the parameter values governing a population’s dynamics. We also modified our terminology throughout the paper to avoid the ambiguity described by the reviewer.

*It is also not clear to me how the authors deal with the multiple (statistical) characteristics of a shift and the various types of shifts in time series data. Studying a single statistical property combined with a rather specific functional form (AICc, Ricker) is not sufficient to declare a change to be a significant (regime) shift, as a shift and specifically a regime shift has several properties (dimensions and shapes) and hence needs to be characterized by more than one (statistical) measure or quantity (see the above papers for this).*

> Thank you for your comment. In light of the revisions to the introduction, we have revised the manuscript to remove the focus specifically on regime shifts, and instead describe our algorithm as a detector for changes to parameter values governing population fluctuations, to avoid this ambiguity.

*The Ricker curve is a simple model used in fisheries to illustrate the S/R relationship (S = stock, R = recruits) as described above. Beside Ricker different other S/R relationships of other shapes and parameterization do exist ((1) initial phase at small stock sizes – (2) intermediate phase at medium stock sizes – (3) final phase at large stock sizes):*

• *Ricker curve shape: (1) positive stock density dependence – (2) maximum/small plateau – (3) negative density dependence*

• *Beverton/Holt curve shape: (1) positive stock density dependence – (2,3) long plateau*

• *Segmented regression shape: (1) positive stock density dependence – (2,3) long plateau.*

*In fisheries sciences the latter is commonly used for break point detection because the two other methods are not suitable for this (given their parameter setup and their shape); apart from this the segmented regression also allows adding further variables quite easily. However, usually applying one of these models to real data in most cases results in a low degree of explanation (poor fits between 10 to 30% only). Hence, 70 to 90% of the variance remains unexplained which makes it difficult to use it for an explicit detection of shifts at small estimation errors (compare also the graphs in the paper which show a huge variation in the data). This leads to high uncertainties in the estimated parameters and consequently huge confidence intervals around them. This clearly hinders the unambiguous or correct recognition of shifts. Moreover, the segmentation as done here does not only reduce the degrees of freedom dramatically (which is a serious constraint for statistical methods because the overall S/R sample sizes are normally relatively small), but also requires to know the number of segments and their size a priori (at least one of these). Furthermore, the Ricker curve is rather specific in its shape and parameter setup with focus on negatively dependent S/R relationships (for instance, cannibalistic or space limited species); it thus hampers a generalization towards other species plus towards all the various other types or forms of shifts (for potential types of shifts see for instance Gröger et al. above).*

>Thank you for your comment. In the paper, we define a dynamic shift not as the movement of a population to different parts of a single curve, as is described here, but a change in parameter values governing the dynamical response, essentially leading to a different Ricker curve with different dynamic rules. We visualize this phenomenon in panel B of figures 3 and 4. Further, we use a general form of the Ricker Model wherein the generational steps of stock/recruitment are not explicitly accounted for; instead, we simply consider population size in year t+1 as a function of the size in year t (described in detail in Peter Turchin’s Complex Population Dynamics 2003, Princeton University Press). Whereas we acknowledge its long history in fisheries research, would argue (as Turchin does) that the Ricker model has much broader applications and it has been applied to many populations beyond fisheries. The Ricker model is also relatively simple and requires limited parameterization, has a non-linear form, and is ‘well behaved’ (i.e. cannot predict populations numbers <0), making it an excellent candidate for our purposes in illustrating the DSD algorithm.

We added text to the introduction to justify our use of this particular population model at the core of the DSD algorithm, which is designed as a means to address problems common to segmented regression: namely, that break points are usually decided arbitrarily, that the models don’t adequately account for nonlinearities, and that uncertainties in the existence and location of breaks are ignored.

We agree with the reviewer that there are limitations in fitting a curve to a single time series and that such approaches can only provide a limited understanding of dynamics. For example, much of the variance in a time series is ignored. However, given data limitations, it is unlikely that all variance will ever be explained- and that which is explained can still be useful to understanding the process under study. Additionally, the dynamic shift detector itself, although implemented with the Ricker model, can be adapted to other model structures- the model selection procedure which underlies its break point location and ranking- is model agnostic. We plan to implement future iterations of the model where the user may specify model structure, however, this was beyond the scope of this initial implementation.

*Finally, in contrast to the common procedure, the authors here add the error term additively, something what is normally done multiplicatively in this context; this leads to a different residual pattern and diagnostics based on a normal instead of a log-normal error distribution. (see for instance,*

*https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4800783/, https://academic.oup.com/icesjms/article/71/8/2307/2804451# , etc.). This clearly affects using the appropriate estimation method (Maximum Likelihood etc.) and its correct application.*

>Thank you for this comment. We simulated error as a function of population size and standard deviation, and then added the resultant term to the prediction, which is similar to how this is handled by the Subbey et al 2014 mentioned above. We now describe this methodology in more detail under the Simulation Study header.

*Simulation and performing scenarios, respectively is usually based on applying specific computational simulation methods (and models). It is not clear to me how the authors did this and whether their simulation method is not in tautological contradiction to their shift detection method (i.e. whether their simulation method and their shift detection method are not of the same computational type or methodological family*

>Thank you for your comment. The simulated data were generated independently of the DSD algorithm, then these data were subjected to the DSD algorithm, and outputs were recorded. Although it is true that the simulated data were generated to have a Ricker structure and are thus not completely random, break points themselves were randomly selected from the set of all possible break points, and we included error in our simulations.