THE ROYAL SOCIETY PUBLISHING
PROCEEDINGS B

Limited evidence for sardine and anchovy asynchrony: re-examining an old story

Margaret C. Siple, Timothy E. Essington, Lewis A. K. Barnett and Mark D. Scheuerell

Article citation details
Proc. R. Soc. B **287**: 20192781.
http://dx.doi.org/10.1098/rspb.2019.2781

Review timeline

Original submission:	27 November 2019
1st revised submission:	3 February 2020
2nd revised submission:	12 February 2020
Final acceptance:	13 February 2020

Note: Reports are unedited and appear as submitted by the referee. The review history appears in chronological order.

Review History

RSPB-2019-2781.R0 (Original submission)

Review form: Reviewer 1 (Geoffrey Hosack)

Recommendation
Major revision is needed (please make suggestions in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Good

General interest: Is the paper of sufficient general interest?
Excellent

Quality of the paper: Is the overall quality of the paper suitable?
Acceptable

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
Yes

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

Is it accessible?
No

Is it clear?
No

Is it adequate?
No

Do you have any ethical concerns with this paper?
No

Comments to the Author
Review of "Limited evidence for sardine and anchovy asynchrony: examining an old story with novel methods"

This paper analyses time series of sardine and anchovy derived from landings data, biomass and recruitment estimates from 5 different regions. The data are analysed for synchrony between sardine and anchovy in two ways. The first method uses a (log) linear state space model (SSM) applied to the time series. Each set of time series (landing and recruitment data, biomass estimates) is analysed independently. The second method uses wavelet analysis of the original time series compared with a summary statistic derived from wavelet analysis applied to a set of simulated data. After the data analysis, a power analysis is developed using simulated data to assess the ability to detect asynchrony using the wavelet approach. The study concludes that covariance between sardine and anchovy is weak for biomass and recruitment time series. Observed asynchrony in landings is determined to be unrelated to biomass. Also, the power to detect low frequency asynchrony from the analysed time series is assessed to be low.

The first set of results (Figure 2) compares the maximum annual landings, estimated biomass and recruitment of anchovy against the corresponding maxima for sardine. It's unclear why dominant species are defined on the basis of the maximum biomass observed in a single year rather than say average or median biomass, which might help smooth annual variation for these species that are known to exhibit high spatio-temporal variability. If these results are retained it would be helpful to provide the reasoning behind the choice of the maxima as the criteria for comparison.

The supporting log linear state space model (SSM) results are in Fig. 3(a), which show the process error covariances. The 95% confidence intervals cover zero for all the time series analysed except for two regions from the recruitment time series. Evidence for asynchrony from SSMs conditioned on the landings data or biomass estimates is described as weak. However, the effect of the non-zero off-diagonal (density dependent) parameters in B are not considered in the discussion of the results. The parameters in the matrix B would also be related to the cross-correlations among species and so should be expected to exhibit joint dependence with the process error covariance when conditioned on the time series data. The density dependence may be important for influencing estimates of negative covariance in the process error among regions, but the density dependent matrix B is not allowed to vary across regions. The reported results across all regions suggest negative asymmetric density dependence is possible for biomass and recruitment, where for each of these two data types one of the 0.95 CIs for the density dependent effect of one species on another excludes zero (Table S4). For landings, the median estimates for the interspecific density dependence appear comparatively weaker than for the other data series. The marginal variances for the process and observation errors are also fixed across all regions. This corresponds to a hypothesis of similar magnitudes of variation in both sampling error and process error across all regions. It is important to discuss the reasoning behind the assumptions of shared parameters among regions and how these might impact inference, for example, of negative temporal dependence between sardine and anchovy. Although the state space model approach would allow for the joint inclusion of landings data, biomass estimates and recruitment estimates into the observation model, this approach was not adopted here. The three data sources may be expected to capture different perspectives of the same population, so why are they analysed independently in this paper? For example, biomass estimates in some cases may take into account the recruitment and landing data (see Barange et al. 2009), which suggests dependence among the time series (for clarity, it might be noted where the biomass estimates are data or may be more appropriately considered model output).

Regarding the log linear state space model analysis, I point out a 2013 study that also used log linear state space models applied to these same landings data from Barange et al. (2009) as used by the authors in this submitted paper, see Hosack, GR, Trenkel, VM, Dambacher, JD. 2013. Journal of Marine Systems 125:77-89 (NB: I've disclosed my identity as reviewer to the authors because I was an author of this paper and so declare this conflict of interest). As might be expected given the close overlap in data and choice of statistical model, there are points of similarities, but also some differences. 1) Hosack et al. (2013) found differences in the marginal process and observation variances for sardine and anchovy among regions. 2) That study showed negative process error correlations among both sardine and anchovy, similar to some of the results presented here in Figure 3a. However, Hosack et al. (2013) found that allowing for cross correlations through the process error covariance matrix alone was better supported by the data rather than via interspecific density-dependence (which would appear in the off-diagonal entries of the B matrix from the authors' study). The more complex case of allowing both interspecific density dependence and correlated process error jointly in the same model, as used in this paper, was not tested. 3) Hosack et al. (2013) examined a 100 year time series of sardine and anchovy, which suggested greater negative correlation for the long time series compared to shorter time series. That result is in line with the reported power analysis from this paper, which suggests analyses of shorter time series may lack power to detect negative temporal dependence among sardine and anchovy. The 2013 paper is pointed out for the three reasons outlined above.

A major conclusion is that the analysis of the landings data, which suggests asynchrony between sardine and anchovy, should be disregarded because the analysis of the biomass estimates suggests otherwise. As noted above, the state space model results based solely on the process error covariance suggested weak dependence, although Table S4 suggests negative density dependence may be possible (and stronger for biomass and recruitment than for landing data). In any case, the conclusion of strong asynchrony in the landings data seems to be drawn from the wavelet analysis results. However, it is acknowledged from the power analysis study that the power to detect low frequency asynchrony is low, which undercuts the conclusion that the study has shown that covariance is weak between sardine and anchovy biomass and recruitment (because this result may follow from the lack of power). For the wavelet analysis methods, the way in which the simulations are constructed is also incompletely described, which obscures the interpretation of results and subsequent conclusions. Nevertheless, the wavelet analysis results are suggestive that landings data, at least at shorter timescales, exhibit stronger negative dependence between sardine and anchovy than either biomass estimates or recruitment data series would suggest. If the wavelet analysis could be better described than this conclusion would be better supported. Additionally, the Discussion notes that separating observation and process error might help interpretation of biomass analyses (lines 379-386). This would also suggest that perhaps it's an SSM analysis, which explicitly represents both process and observation error in a parametric model, that could be considered as the primary analysis to base the study conclusions on.

Specific comments:

Lines 105-106: The cited article Izquierdo-Pena et al. (2018) distinguishes 1) synchrony between populations of different regions and 2) alternations between species within regions. In this introductory paragraph, both interpretations (1) and (2) are possible for “asynchrony”. The authors should clarify here what is meant by “asynchrony” in their paper: Is it (1) or (2) or both?

Lines 117-119: can remove parentheses here for the cape gannet example

Line 123: “ … depends on life history traits and strategies”. The distinction between integration over space and time should be made clear. Provide guidance on how integration over time depends on life history traits and strategies in a way that is different from integration over space.

Line 133-136: run-on sentence

Lines 148-150: “… surprisingly few attempts …” Although, as suggested by the title, there have been previous attempts to estimate alternation of species with regions and also synchrony of populations among regions. Some of these studies are already cited elsewhere in this paper. One option is to omit “few” here and cite the studies that have looked at these questions previously.

Lines 154-155: “… we analyse the dynamics of the populations themselves …” This sentence is contrasting the current study with the analysis of landings data by Izquierdo-Pena et al. (2018) who argue that landings provide a proxy indicator of abundance or biomass. The suggestion is that the current study has access to alternative data such that the analysis is of the actual populations. The reasoning behind this suggestion should be explained here (later in the paper, the Discussion does note that the biomass time series may not directly correspond to true population abundance).

Line 184: “… most comprehensive study of trends so far …” This refers again to Izquierdo-Pena et al. (2018), however, there are studies that similarly have looked at historical landings including time series of longer length. It is ambiguous how this citation is the most comprehensive study. One option is to define what is meant by “most comprehensive”, another is to cite the related studies together.

Line 190-191: Why define dominant species based on maximum biomass rather than say average or median biomass?

Lines 202-203: “… we fit a MARSS model pairs of sardine and anchovy time series …” Clarify that the analysis is applied to sardine and anchovy time series in each region. What are the assumptions of independence among regions? Also, it would be helpful to describe which parameters are assumed identical among regions and why these choices were made. Similarly, describe how the different data series (landings, biomass estimates, recruit estimates) are to be analysed, e.g., jointly or separately. How might the three data series provide different perspectives on the same population?

Lines 207-209: “ … without phase information.” Insufficient information is provided on how the surrogate simulations were constructed. The citation to the fractal R package is not enough for the simulation procedure to be understood or reproduced.

Lines 222-224: “… somewhat unusual way of looking at wavelet information …” The analysis ignores how values are arranged in the time domain because of the binning by frequency, so why not use Fourier (cross-spectrum) analysis?

Line 233: The covariance matrix for the process errors should be denoted using a capital letter (because it’s a matrix).

Line 234--235: Why use “m x 1 vector” and “m x m” matrix when C is a 2 x 1 vector? It must be that m = 2.

Lines 248-249: “Because we were interested in the detectability … we did not simulate observation error. “ If there is an interest in detectability, isn’t this an argument in favour of including observation error in the simulations?

Lines 283-288: “Covariance in biomass was similarly weak …” The results focus on the process error covariance matrix but the discussion neglects the non-zero off-diagonal entries of the matrix B that are also related to the cross-correlations among the sardine and anchovy time series within each region and hence likely have joint dependence with the process error covariance matrix estimates. Also Table S4 suggests that some negative dependence could be capture by these off-diagonal entries of B.

Line 288-230: The range provided for the intraspecific density dependence terms does not include one (equal to density independence). Why is density dependence said to be weak?

Lines 303-304: “Across all ecosystems, WMR distributions indicated stronger asynchrony in landings …” Fig. 3b suggests it’s unclear that this is true for timescales greater than 10 years in the Japan region.

Lines 313-314: If neither the methods nor the results of an analysis are presented then it is probably better not to report its conclusions (some journals do not allow it). Also remove reference to the analysis in the Discussion. Or add methods/results from the correlation analysis to paper.

Lines 328-332: “… we show that … taxa are not strongly asynchronous … observations of asynchrony in landings are not indicative of underlying patterns in biomass or recruitment.” It’s unclear that this result was in fact shown. Note that line 308 says that “statistical power to detect asynchrony at decadal scales was low”. The reported process covariance terms do not account for the off-diagonal entries of B.

Lines 337-338: “Significant asynchrony … only occurred in the Northeast Atlantic and Humboldt …” What result(s) is this sentence referring to?

Lines 342-343: “In the case of the Benguela Current, our finding of synchrony and positive covariance is in agreement with the findings of Izquierdo-Peña et al. [27].” On the contrary, the latter study says that “sardines and anchovies from the Benguela showed alternation in the last part of the last Century” though “In Benguela, alternation between species is not evident during at least the last two decades, as both species have remained at low productivity levels.”

Lines 383-385: “If observation error is low … it should be possible to improve detection with better data”. In general, wouldn’t collecting better data reduce observation error?

Eqs. S1 and S2: Why are the marginal process error variances are held constant across the regions whereas the process error correlation is allowed to change? Why are observation errors held constant across regions? These hypotheses suggest similar levels of environmental stochasticity and observation error for all regions considered, and only correlation in the process error can vary across regions. Also B is held constant across regions. What is the reasoning for these assumptions?

Eq. S5: It’s unclear which parameters vary across regions and which are held fixed.

Eq. S6: It’s unclear why the observation error covariance matrix R should depend on B.

Last equation and above: equation numbering is out of sequence with references in the captions.

Review form: Reviewer 2

Recommendation
Major revision is needed (please make suggestions in comments)

Scientific importance: Is the manuscript an original and important contribution to its field?
Acceptable

General interest: Is the paper of sufficient general interest?
Acceptable

Quality of the paper: Is the overall quality of the paper suitable?
Acceptable

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
No

Do you have any concerns about statistical analyses in this paper? If so, please specify them explicitly in your report.
Yes

It is a condition of publication that authors make their supporting data, code and materials available - either as supplementary material or hosted in an external repository. Please rate, if applicable, the supporting data on the following criteria.

> **Is it accessible?**
> Yes
>
> **Is it clear?**
> Yes
>
> **Is it adequate?**
> Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
The authors analyzed five stocks of sardine and anchovy for comparability and asynchrony. They also performed a power analysis using a simulation study. Overall I agree with the authors that previous studies have probably exaggerated the degree of asynchrony in sardines and anchovy given the relatively short time series that are typically used. However the authors could do a better job of justifying their methodological choices. Particularly their definition and analysis of comparability was not well justified.

Title - The methods are not really novel since both the models and data have been used before in different studies. Perhaps a re-analysis of an old story is more accurate.

Abstract - Both the abstract and introduction should state the number of stocks that were analyzed in this study (five?).

Introduction - A sentence or two should be added to the end of the introduction that mentions the data and models that will be used (MARSS, wavelet, etc.). It seems like the first paragraph of the methods Data section should actually be the last paragraph of the Introduction since it does not really describe the data but rather describes the entire study approach.

Methods-
Line 191- Why is the log-ratio of maximum landings and biomass a good metric of comparability? Why not choose the mean or median biomass which will be more stable? The maximum seems arbitrary and not particularly ecologically relevant. If this how it has been defined in the past then that should be referenced.
Line 206 and 252- 50 surrogates seems like a small sample. I would expect hundreds or more. Why just 50? A sensitivity analysis to this would be nice for the supplement.
Line 222- If it is unusual then what is usual and why was that not done?
Line 248 - You have estimates of observation error from your MARSS fits to the real data so why not use them in the simulations? Aren't you are interested in the detectability of asynchrony given time series length and realistic observation error? Or perhaps you can state that this is a conservative estimate of time series length since in reality we will have noisy observations which would make the required length even longer.

Results-
What is the null hypothesis for the comparability analysis? We wouldn't expect the maximums to be exactly equal, so saying they're different doesn't seem very meaningful. If the uncertainty of the ratio could be quantified (perhaps using the mean rather than the maximum) then you could perform a hypothesis test of whether the mean is significantly different from some null.

Discussion -
Perhaps mention whether you think asynchrony could be detected in the longer geological scale-deposition based time series (e.g., Holmgren-Urba & Baumgartner 1993).

Decision letter (RSPB-2019-2781.R0)

13-Jan-2020

Dear Dr Siple:

Your manuscript has now been peer reviewed and the reviews have been assessed by an Associate Editor. The reviewers' comments (not including confidential comments to the Editor) and the comments from the Associate Editor are included at the end of this email for your reference. As you will see, the reviewers and the Editors have raised some concerns with your manuscript and we would like to invite you to revise your manuscript to address them.

We do not allow multiple rounds of revision so we urge you to make every effort to fully address all of the comments at this stage. If deemed necessary by the Associate Editor, your manuscript will be sent back to one or more of the original reviewers for assessment. If the original reviewers are not available we may invite new reviewers. Please note that we cannot guarantee eventual acceptance of your manuscript at this stage.

To submit your revision please log into http://mc.manuscriptcentral.com/prsb and enter your

Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions", click on "Create a Revision". Your manuscript number has been appended to denote a revision.

When submitting your revision please upload a file under "Response to Referees" - in the "File Upload" section. This should document, point by point, how you have responded to the reviewers' and Editors' comments, and the adjustments you have made to the manuscript. We require a copy of the manuscript with revisions made since the previous version marked as 'tracked changes' to be included in the 'response to referees' document.

Your main manuscript should be submitted as a text file (doc, txt, rtf or tex), not a PDF. Your figures should be submitted as separate files and not included within the main manuscript file.

When revising your manuscript you should also ensure that it adheres to our editorial policies (https://royalsociety.org/journals/ethics-policies/). You should pay particular attention to the following:

Research ethics:
If your study contains research on humans please ensure that you detail in the methods section whether you obtained ethical approval from your local research ethics committee and gained informed consent to participate from each of the participants.

Use of animals and field studies:
If your study uses animals please include details in the methods section of any approval and licences given to carry out the study and include full details of how animal welfare standards were ensured. Field studies should be conducted in accordance with local legislation; please include details of the appropriate permission and licences that you obtained to carry out the field work.

Data accessibility and data citation:
It is a condition of publication that you make available the data and research materials supporting the results in the article. Datasets should be deposited in an appropriate publicly available repository and details of the associated accession number, link or DOI to the datasets must be included in the Data Accessibility section of the article (https://royalsociety.org/journals/ethics-policies/data-sharing-mining/). Reference(s) to datasets should also be included in the reference list of the article with DOIs (where available).

In order to ensure effective and robust dissemination and appropriate credit to authors the dataset(s) used should also be fully cited and listed in the references.

If you wish to submit your data to Dryad (http://datadryad.org/) and have not already done so you can submit your data via this link http://datadryad.org/submit?journalID=RSPB&manu=(Document not available), which will take you to your unique entry in the Dryad repository.

If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link.

For more information please see our open data policy http://royalsocietypublishing.org/data-sharing.

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the

accompanying article so that the supplementary material can be attributed a unique DOI. Please try to submit all supplementary material as a single file.

Online supplementary material will also carry the title and description provided during submission, so please ensure these are accurate and informative. Note that the Royal Society will not edit or typeset supplementary material and it will be hosted as provided. Please ensure that the supplementary material includes the paper details (authors, title, journal name, article DOI). Your article DOI will be 10.1098/rspb.[paper ID in form xxxx.xxxx e.g. 10.1098/rspb.2016.0049].

Please submit a copy of your revised paper within three weeks. If we do not hear from you within this time your manuscript will be rejected. If you are unable to meet this deadline please let us know as soon as possible, as we may be able to grant a short extension.

Thank you for submitting your manuscript to Proceedings B; we look forward to receiving your revision. If you have any questions at all, please do not hesitate to get in touch.

Best wishes,
Dr Daniel Costa
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
Both reviewers have recommended that major revisions are required in order to bring this manuscript up to publishable quality, with concerns raised around the analytical approach undertaken and the interpretation of the associated results. Both have provided detailed comments on the manuscript and suggestions for revising the manuscript that should be considered carefully by the authors.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
Review of “Limited evidence for sardine and anchovy asynchrony: examining an old story with novel methods”

This paper analyses time series of sardine and anchovy derived from landings data, biomass and recruitment estimates from 5 different regions. The data are analysed for synchrony between sardine and anchovy in two ways. The first method uses a (log) linear state space model (SSM) applied to the time series. Each set of time series (landing and recruitment data, biomass estimates) is analysed independently. The second method uses wavelet analysis of the original time series compared with a summary statistic derived from wavelet analysis applied to a set of simulated data. After the data analysis, a power analysis is developed using simulated data to assess the ability to detect asynchrony using the wavelet approach. The study concludes that covariance between sardine and anchovy is weak for biomass and recruitment time series. Observed asynchrony in landings is determined to be unrelated to biomass. Also, the power to detect low frequency asynchrony from the analysed time series is assessed to be low.

The first set of results (Figure 2) compares the maximum annual landings, estimated biomass and recruitment of anchovy against the corresponding maxima for sardine. It’s unclear why dominant species are defined on the basis of the maximum biomass observed in a single year rather than say average or median biomass, which might help smooth annual variation for these species that are known to exhibit high spatio-temporal variability. If these results are retained it would be helpful to provide the reasoning behind the choice of the maxima as the criteria for comparison.

The supporting log linear state space model (SSM) results are in Fig. 3(a), which show the process error covariances. The 95% confidence intervals cover zero for all the time series analysed except for two regions from the recruitment time series. Evidence for asynchrony from SSMs conditioned on the landings data or biomass estimates is described as weak. However, the effect of the non-zero off-diagonal (density dependent) parameters in B are not considered in the discussion of the results. The parameters in the matrix B would also be related to the cross-correlations among species and so should be expected to exhibit joint dependence with the process error covariance when conditioned on the time series data. The density dependence may be important for influencing estimates of negative covariance in the process error among regions, but the density dependent matrix B is not allowed to vary across regions. The reported results across all regions suggest negative asymmetric density dependence is possible for biomass and recruitment, where for each of these two data types one of the 0.95 CIs for the density dependent effect of one species on another excludes zero (Table S4). For landings, the median estimates for the interspecific density dependence appear comparatively weaker than for the other data series. The marginal variances for the process and observation errors are also fixed across all regions. This corresponds to a hypothesis of similar magnitudes of variation in both sampling error and process error across all regions. It is important to discuss the reasoning behind the assumptions of shared parameters among regions and how these might impact inference, for example, of negative temporal dependence between sardine and anchovy. Although the state space model approach would allow for the joint inclusion of landings data, biomass estimates and recruitment estimates into the observation model, this approach was not adopted here. The three data sources may be expected to capture different perspectives of the same population, so why are they analysed independently in this paper? For example, biomass estimates in some cases may take into account the recruitment and landing data (see Barange et al. 2009), which suggests dependence among the time series (for clarity, it might be noted where the biomass estimates are data or may be more appropriately considered model output).

Regarding the log linear state space model analysis, I point out a 2013 study that also used log linear state space models applied to these same landings data from Barange et al. (2009) as used by the authors in this submitted paper, see Hosack, GR, Trenkel, VM, Dambacher, JD. 2013. Journal of Marine Systems 125:77-89 (NB: I've disclosed my identity as reviewer to the authors because I was an author of this paper and so declare this conflict of interest). As might be expected given the close overlap in data and choice of statistical model, there are points of similarities, but also some differences. 1) Hosack et al. (2013) found differences in the marginal process and observation variances for sardine and anchovy among regions. 2) That study showed negative process error correlations among both sardine and anchovy, similar to some of the results presented here in Figure 3a. However, Hosack et al. (2013) found that allowing for cross correlations through the process error covariance matrix alone was better supported by the data rather than via interspecific density-dependence (which would appear in the off-diagonal entries of the B matrix from the authors' study). The more complex case of allowing both interspecific density dependence and correlated process error jointly in the same model, as used in this paper, was not tested. 3) Hosack et al. (2013) examined a 100 year time series of sardine and anchovy, which suggested greater negative correlation for the long time series compared to shorter time series. That result is in line with the reported power analysis from this paper, which suggests analyses of shorter time series may lack power to detect negative temporal dependence among sardine and anchovy. The 2013 paper is pointed out for the three reasons outlined above.

A major conclusion is that the analysis of the landings data, which suggests asynchrony between sardine and anchovy, should be disregarded because the analysis of the biomass estimates suggests otherwise. As noted above, the state space model results based solely on the process error covariance suggested weak dependence, although Table S4 suggests negative density dependence may be possible (and stronger for biomass and recruitment than for landing data). In any case, the conclusion of strong asynchrony in the landings data seems to be drawn from the wavelet analysis results. However, it is acknowledged from the power analysis study that the power to detect low frequency asynchrony is low, which undercuts the conclusion that the study has shown that covariance is weak between sardine and anchovy biomass and recruitment (because this result may follow from the lack of power). For the wavelet analysis methods, the way in which the simulations are constructed is also incompletely described, which obscures the interpretation of results and subsequent conclusions. Nevertheless, the wavelet analysis results are suggestive that landings data, at least at shorter timescales, exhibit stronger negative dependence between sardine and anchovy than either biomass estimates or recruitment data series would suggest. If the wavelet analysis could be better described than this conclusion would be better supported. Additionally, the Discussion notes that separating observation and process error might help interpretation of biomass analyses (lines 379-386). This would also suggest that perhaps it's an SSM analysis, which explicitly represents both process and observation error in a parametric model, that could be considered as the primary analysis to base the study conclusions on.

Specific comments:

Lines 105-106: The cited article Izquierdo-Pena et al. (2018) distinguishes 1) synchrony between populations of different regions and 2) alternations between species within regions. In this introductory paragraph, both interpretations (1) and (2) are possible for "asynchrony". The authors should clarify here what is meant by "asynchrony" in their paper: Is it (1) or (2) or both?

Lines 117-119: can remove parentheses here for the cape gannet example

Line 123: " … depends on life history traits and strategies". The distinction between integration over space and time should be made clear. Provide guidance on how integration over time depends on life history traits and strategies in a way that is different from integration over space.

Line 133-136: run-on sentence

Lines 148-150: "… surprisingly few attempts …" Although, as suggested by the title, there have been previous attempts to estimate alternation of species with regions and also synchrony of populations among regions. Some of these studies are already cited elsewhere in this paper. One option is to omit "few" here and cite the studies that have looked at these questions previously.

Lines 154-155: "… we analyse the dynamics of the populations themselves …" This sentence is contrasting the current study with the analysis of landings data by Izquierdo-Pena et al. (2018) who argue that landings provide a proxy indicator of abundance or biomass. The suggestion is that the current study has access to alternative data such that the analysis is of the actual populations. The reasoning behind this suggestion should be explained here (later in the paper, the Discussion does note that the biomass time series may not directly correspond to true population abundance).

Line 184: "… most comprehensive study of trends so far …" This refers again to Izquierdo-Pena et al. (2018), however, there are studies that similarly have looked at historical landings including time series of longer length. It is ambiguous how this citation is the most comprehensive study. One option is to define what is meant by "most comprehensive", another is to cite the related studies together.

Line 190-191: Why define dominant species based on maximum biomass rather than say average or median biomass?

Lines 202-203: "… we fit a MARSS model pairs of sardine and anchovy time series …" Clarify that the analysis is applied to sardine and anchovy time series in each region. What are the assumptions of independence among regions? Also, it would be helpful to describe which parameters are assumed identical among regions and why these choices were made. Similarly, describe how the different data series (landings, biomass estimates, recruit estimates) are to be analysed, e.g., jointly or separately. How might the three data series provide different perspectives on the same population?

Lines 207-209: " … without phase information." Insufficient information is provided on how the surrogate simulations were constructed. The citation to the fractal R package is not enough for the simulation procedure to be understood or reproduced.

Lines 222-224: "… somewhat unusual way of looking at wavelet information …" The analysis ignores how values are arranged in the time domain because of the binning by frequency, so why not use Fourier (cross-spectrum) analysis?

Line 233: The covariance matrix for the process errors should be denoted using a capital letter (because it's a matrix).

Line 234--235: Why use "m x 1 vector" and "m x m" matrix when C is a 2 x 1 vector? It must be that m = 2.

Lines 248-249: "Because we were interested in the detectability … we did not simulate observation error. " If there is an interest in detectability, isn't this an argument in favour of including observation error in the simulations?

Lines 283-288: "Covariance in biomass was similarly weak …" The results focus on the process error covariance matrix but the discussion neglects the non-zero off-diagonal entries of the matrix B that are also related to the cross-correlations among the sardine and anchovy time series within each region and hence likely have joint dependence with the process error covariance matrix estimates. Also Table S4 suggests that some negative dependence could be capture by these off-diagonal entries of B.

Line 288-230: The range provided for the intraspecific density dependence terms does not include one (equal to density independence). Why is density dependence said to be weak?

Lines 303-304: "Across all ecosystems, WMR distributions indicated stronger asynchrony in landings …" Fig. 3b suggests it's unclear that this is true for timescales greater than 10 years in the Japan region.

Lines 313-314: If neither the methods nor the results of an analysis are presented then it is probably better not to report its conclusions (some journals do not allow it). Also remove reference to the analysis in the Discussion. Or add methods/results from the correlation analysis to paper.

Lines 328-332: "… we show that … taxa are not strongly asynchronous … observations of asynchrony in landings are not indicative of underlying patterns in biomass or recruitment." It's unclear that this result was in fact shown. Note that line 308 says that "statistical power to detect asynchrony at decadal scales was low". The reported process covariance terms do not account for the off-diagonal entries of B.

Lines 337-338: "Significant asynchrony … only occurred in the Northeast Atlantic and Humboldt …" What result(s) is this sentence referring to?

Lines 342-343: "In the case of the Benguela Current, our finding of synchrony and positive covariance is in agreement with the findings of Izquierdo-Peña et al. [27]." On the contrary, the latter study says that "sardines and anchovies from the Benguela showed alternation in the last part of the last Century" though "In Benguela, alternation between species is not evident during at least the last two decades, as both species have remained at low productivity levels."

Lines 383-385: “If observation error is low … it should be possible to improve detection with better data”. In general, wouldn’t collecting better data reduce observation error?

Eqs. S1 and S2: Why are the marginal process error variances are held constant across the regions whereas the process error correlation is allowed to change? Why are observation errors held constant across regions? These hypotheses suggest similar levels of environmental stochasticity and observation error for all regions considered, and only correlation in the process error can vary across regions. Also B is held constant across regions. What is the reasoning for these assumptions?

Eq. S5: It’s unclear which parameters vary across regions and which are held fixed.

Eq. S6: It’s unclear why the observation error covariance matrix R should depend on B.

Last equation and above: equation numbering is out of sequence with references in the captions.

Referee: 2

Comments to the Author(s)

The authors analyzed five stocks of sardine and anchovy for comparability and asynchrony. They also performed a power analysis using a simulation study. Overall I agree with the authors that previous studies have probably exaggerated the degree of asynchrony in sardines and anchovy given the relatively short time series that are typically used. However the authors could do a better job of justifying their methodological choices. Particularly their definition and analysis of comparability was not well justified.

Title - The methods are not really novel since both the models and data have been used before in different studies. Perhaps a re-analysis of an old story is more accurate.

Abstract - Both the abstract and introduction should state the number of stocks that were analyzed in this study (five?).

Introduction - A sentence or two should be added to the end of the introduction that mentions the data and models that will be used (MARSS, wavelet, etc.). It seems like the first paragraph of the methods Data section should actually be the last paragraph of the Introduction since it does not really describe the data but rather describes the entire study approach.

Methods-
Line 191- Why is the log-ratio of maximum landings and biomass a good metric of comparability? Why not choose the mean or median biomass which will be more stable? The maximum seems arbitrary and not particularly ecologically relevant. If this how it has been defined in the past then that should be referenced.
Line 206 and 252- 50 surrogates seems like a small sample. I would expect hundreds or more. Why just 50? A sensitivity analysis to this would be nice for the supplement.
Line 222- If it is unusual then what is usual and why was that not done?
Line 248 - You have estimates of observation error from your MARSS fits to the real data so why not use them in the simulations? Aren't you are interested in the detectability of asynchrony given time series length and realistic observation error? Or perhaps you can state that this is a conservative estimate of time series length since in reality we will have noisy observations which would make the required length even longer.

Results-
What is the null hypothesis for the comparability analysis? We wouldn't expect the maximums to be exactly equal, so saying they're different doesn't seem very meaningful. If the uncertainty of the ratio could be quantified (perhaps using the mean rather than the maximum) then you could perform a hypothesis test of whether the mean is significantly different from some null.

Discussion -
Perhaps mention whether you think asynchrony could be detected in the longer geological scale-deposition based time series (e.g., Holmgren-Urba & Baumgartner 1993).

Author's Response to Decision Letter for (RSPB-2019-2781.R0)

See Appendix A.

Decision letter (RSPB-2019-2781.R1)

06-Feb-2020

Dear Dr Siple

I am pleased to inform you that your Review manuscript RSPB-2019-2781.R1 entitled "Limited evidence for sardine and anchovy asynchrony: re-examining an old story" has been accepted for publication in Proceedings B.

The referee(s) do not recommend any further changes. Therefore, please proof-read your manuscript carefully and upload your final files for publication. Because the schedule for publication is very tight, it is a condition of publication that you submit the revised version of your manuscript within 7 days. If you do not think you will be able to meet this date please let me know immediately.

To upload your manuscript, log into http://mc.manuscriptcentral.com/prsb and enter your Author Centre, where you will find your manuscript title listed under "Manuscripts with Decisions." Under "Actions," click on "Create a Revision." Your manuscript number has been appended to denote a revision.

You will be unable to make your revisions on the originally submitted version of the manuscript. Instead, upload a new version through your Author Centre.

Before uploading your revised files please make sure that you have:

1) A text file of the manuscript (doc, txt, rtf or tex), including the references, tables (including captions) and figure captions. Please remove any tracked changes from the text before submission. PDF files are not an accepted format for the "Main Document".

2) A separate electronic file of each figure (tiff, EPS or print-quality PDF preferred). The format should be produced directly from original creation package, or original software format. Please note that PowerPoint files are not accepted.

3) Electronic supplementary material: this should be contained in a separate file from the main text and the file name should contain the author’s name and journal name, e.g authorname_procb_ESM_figures.pdf

All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI. Please see: https://royalsociety.org/journals/authors/author-guidelines/

4) Data-Sharing and data citation
It is a condition of publication that data supporting your paper are made available. Data should be made available either in the electronic supplementary material or through an appropriate repository. Details of how to access data should be included in your paper. Please see https://royalsociety.org/journals/ethics-policies/data-sharing-mining/ for more details.

If you wish to submit your data to Dryad (http://datadryad.org/) and have not already done so you can submit your data via this link
http://datadryad.org/submit?journalID=RSPB&manu=RSPB-2019-2781.R1 which will take you to your unique entry in the Dryad repository.

If you have already submitted your data to dryad you can make any necessary revisions to your dataset by following the above link.

5) For more information on our Licence to Publish, Open Access, Cover images and Media summaries, please visit https://royalsociety.org/journals/authors/author-guidelines/.

Once again, thank you for submitting your manuscript to Proceedings B and I look forward to receiving your final version. If you have any questions at all, please do not hesitate to get in touch.

Sincerely,
Dr Daniel Costa
Editor, Proceedings B
mailto:proceedingsb@royalsociety.org

Associate Editor Board Member
Comments to Author:
The authors have made revisions to the paper in line with the reviewers comments. Note that the references haven't updated so the authors will need to provide these

Decision letter (RSPB-2019-2781.R2)

13-Feb-2020

Dear Dr Siple

I am pleased to inform you that your manuscript entitled "Limited evidence for sardine and anchovy asynchrony: re-examining an old story" has been accepted for publication in Proceedings B.

You can expect to receive a proof of your article from our Production office in due course, please check your spam filter if you do not receive it. PLEASE NOTE: you will be given the exact page length of your paper which may be different from the estimation from Editorial and you may be asked to reduce your paper if it goes over the 10 page limit.

If you are likely to be away from e-mail contact please let us know. Due to rapid publication and an extremely tight schedule, if comments are not received, we may publish the paper as it stands.

If you have any queries regarding the production of your final article or the publication date please contact procb_proofs@royalsociety.org

Your article has been estimated as being 10 pages long. Our Production Office will be able to confirm the exact length at proof stage.

Open Access
You are invited to opt for Open Access, making your freely available to all as soon as it is ready for publication under a CCBY licence. Our article processing charge for Open Access is £1700. Corresponding authors from member institutions (http://royalsocietypublishing.org/site/librarians/allmembers.xhtml) receive a 25% discount to these charges. For more information please visit http://royalsocietypublishing.org/open-access.

Paper charges
An e-mail request for payment of any related charges will be sent out shortly. The preferred payment method is by credit card; however, other payment options are available.

Electronic supplementary material:
All supplementary materials accompanying an accepted article will be treated as in their final form. They will be published alongside the paper on the journal website and posted on the online figshare repository. Files on figshare will be made available approximately one week before the accompanying article so that the supplementary material can be attributed a unique DOI.

You are allowed to post any version of your manuscript on a personal website, repository or preprint server. However, the work remains under media embargo and you should not discuss it with the press until the date of publication. Please visit https://royalsociety.org/journals/ethics-policies/media-embargo for more information.

Thank you for your fine contribution. On behalf of the Editors of the Proceedings B, we look forward to your continued contributions to the Journal.

Sincerely,
Proceedings B
mailto: proceedingsb@royalsociety.org

Appendix A

Manuscript ID RSPB-2019-2781

I. **Responses to reviewer comments**

Referee 1

1.1 The first set of results (Figure 2) compares the maximum annual landings, estimated biomass and recruitment of anchovy against the corresponding maxima for sardine. It's unclear why dominant species are defined on the basis of the maximum biomass observed in a single year rather than say average or median biomass, which might help smooth annual variation for these species that are known to exhibit high spatio-temporal variability. If these results are retained it would be helpful to provide the reasoning behind the choice of the maxima as the criteria for comparison.

We have changed Figure 2 to show the log-ratio of the median biomasses for sardine vs. anchovy, and now define the "dominant" stock (between sardine and anchovy) as the species with the highest median biomass (instead of highest maximum) (Methods, L195-196; Results, L288-292).

This has changed our original result that landings were more comparable between the two species: now, spawning stock biomass indicates more comparability (i.e., the median spawning stock biomasses are more similar than median landings). We interpret this to mean that when one species is dominant in terms of biomass, the landings are strongly dominated by that species. This change in result aligns more closely with our conclusion that landings are exaggerating natural patterns, as they do with asynchrony.

1.2 The supporting log linear state space model (SSM) results are in Fig. 3(a), which show the process error covariances. The 95% confidence intervals cover zero for all the time series analysed except for two regions from the recruitment time series. Evidence for asynchrony from SSMs conditioned on the landings data or biomass estimates is described as weak. However, the effect of the non-zero off-diagonal (density dependent) parameters in B are not considered in the discussion of the results. The parameters in the matrix B would also be related to the cross-correlations among species and so should be expected to exhibit joint dependence with the process error covariance when conditioned on the time series data. The density dependence may be important for influencing estimates of negative covariance in the process error among regions, but the density dependent matrix B is not allowed to vary across regions.

The covariance between the two species is indeed a function of both the **B** and **Q** matrices. Originally, we assumed that the **B** matrix was identical (no interactions among the values of x), which I think is also the way that Hosack et al. (2013) set up their state-space model. The **B** matrix estimation was added after a previous reviewer recommended it, in order to explicitly characterize species interactions, and the form of the B matrix we chose (same autocorrelation and interspecific effects across regions) had the lowest AICc of the candidate models (all other options were more complex forms of **B**). This reviewer's questions and comments about the **B** matrix have given us the opportunity to revisit the model and we appreciate this opportunity. The reasons we have decided to estimate a global value of **B** are:

- The model does not converge when we try to estimate a unique **B** and **Q** matrix for each region- this is not surprising because **B** and **Q** both relate to interactions between the elements of **x**.

- Estimating a shared **B** among regions characterizes sardine-anchovy interactions that occur independently from the environment (while the **Q** matrix characterizes environmental effects alone), and this allows us to incorporate a realistic amount of autocorrelation and species interactions in the power analysis.
- Finally, calculating $\det(B)^2$ can provide the proportion of total variance in x_t owing to species interactions, which is a useful piece of information.

We have modified the text in the following places to be more clear about why we chose to structure the model this way:

- In the supplementary methods: we have added the above reasoning when we explain why we estimated covariance the way we did (ESM)
- In the text (L 255-259): We specify that we assumed direct interactions (e.g., what one would find in an isolated experiment with the two species) were similar across ecosystems but variance/covariance was estimated as independent: "To estimate covariance between sardine and anchovy, we fit a MARSS model to the dominant sardine and anchovy stock in each region, using the MARSS package in R [1,2]. We assumed that species interactions (independent of the environment) were similar across ecosystems, but that their covariance was region-specific, an assumption supported by model fits (see ESM for full model structure)."

We are limited by Proceedings B's space requirements in the main document, so the more detailed description is in the ESM.

1.3 The reported results across all regions suggest negative asymmetric density dependence is possible for biomass and recruitment, where for each of these two data types one of the 0.95 CIs for the density dependent effect of one species on another excludes zero (Table S4). For landings, the median estimates for the interspecific density dependence appear comparatively weaker than for the other data series. The marginal variances for the process and observation errors are also fixed across all regions. This corresponds to a hypothesis of similar magnitudes of variation in both sampling error and process error across all regions. It is important to discuss the reasoning behind the assumptions of shared parameters among regions and how these might impact inference, for example, of negative temporal dependence between sardine and anchovy.
We did make a simplifying assumption that observation error was similar across all regions. We made this assumption because a) we don't have any evidence or logic to support why or how the observation errors for biomass estimates would differ substantially across regions, b) we cannot partition observation and process errors from the stock assessment outputs, let alone estimate each of them separately for each region, and c) because many of the issues with detecting changes in biomass are shared across stocks because of their similar biological characteristics (e.g., hyperstability because of aggregative behavior, or higher interannual variability because of changes in predator abundance and thus changes in natural mortality).

1.4 Although the state space model approach would allow for the joint inclusion of landings data, biomass estimates and recruitment estimates into the observation model, this approach was not adopted here. The three data sources may be expected to capture different perspectives of the same population, so why are they analysed independently in this paper? For example, biomass estimates in some cases may take into account the recruitment and landing data (see Barange et al. 2009), which suggests dependence among the time series (for clarity, it might be noted where the biomass estimates are data or may be more appropriately considered model output).

For all the biomass time series in this study, the values were estimates from stock assessment models. This is an important point and have now included it in the Methods (section 2a; L192-193).

1.5 Regarding the log linear state space model analysis, I point out a 2013 study that also used log linear state space models applied to these same landings data from Barange et al. (2009) as used by the authors in this submitted paper, see Hosack, GR, Trenkel, VM, Dambacher, JD. 2013. Journal of Marine Systems 125:77-89 (NB: I've disclosed my identity as reviewer to the authors because I was an author of this paper and so declare this conflict of interest). As might be expected given the close overlap in data and choice of statistical model, there are points of similarities, but also some differences.
1) Hosack et al. (2013) found differences in the marginal process and observation variances for sardine and anchovy among regions.
We very much appreciate the reviewer bringing our attention to this paper. We did not to estimate observation error for spawning stock biomass and recruitment, because biomass and recruitment are virtually all estimates already from stock assessment models (and thus might be considered "states")—and we apply this same approach to landings although it would be suitable to treat landings as observations instead of states. We clarify which time series are data vs. estimates in the Methods (L223-229) and describe in the supplement our decision to treat both landings and biomass as independent data (first section of ESM, "estimating covariance"). We also provide Hosack et al. as an example of how observation vs. process error could be partitioned (L620).

1.6
2) That study showed negative process error correlations among both sardine and anchovy, similar to some of the results presented here in Figure 3a. However, Hosack et al. (2013) found that allowing for cross correlations through the process error covariance matrix alone was better supported by the data rather than via interspecific density-dependence (which would appear in the off-diagonal entries of the B matrix from the authors' study). The more complex case of allowing both interspecific density dependence and correlated process error jointly in the same model, as used in this paper, was not tested.
We appreciate the reviewer's approach- in fact, the **B** matrix estimation was included in our submission at the suggestion of a previous reviewer—so our original analysis also estimated **Q** alone, and we think this approach is quite sufficient. We decided on estimating **B** because we wanted to have a way to explicitly characterize species interactions independent of environmental variability. We found that the estimates of **Q** did not change noticeably when **B** was estimated vs. fixed, and we suspect that is because the two matrices both characterize covariance in different ways. We have added material to the methods that explains the assumptions behind the way we structured the model (L255-259) and have added material in the results and the supplement that explain the relationship between **B** and **Q**, and provide code for readers who are interested in experimenting with the relationship.

1.7 Hosack et al. (2013) examined a 100 year time series of sardine and anchovy, which suggested greater negative correlation for the long time series compared to shorter time series. That result is in line with the reported power analysis from this paper, which suggests analyses of shorter time series may lack power to detect negative temporal dependence among sardine and

anchovy. The 2013 paper is pointed out for the three reasons outlined above.
We now mention Hosack et al.'s result in the Discussion (L 419-421). We are pleased that our findings are similar. We think it draws a useful contrast to Baumgartner et al. (1992), a very long time series in which the authors also concluded that asynchrony was not a dominant pattern.

1.8 A major conclusion is that the analysis of the landings data, which suggests asynchrony between sardine and anchovy, should be disregarded because the analysis of the biomass estimates suggests otherwise. As noted above, the state space model results based solely on the process error covariance suggested weak dependence, although Table S4 suggests negative density dependence may be possible (and stronger for biomass and recruitment than for landing data). In any case, the conclusion of strong asynchrony in the landings data seems to be drawn from the wavelet analysis results. However, it is acknowledged from the power analysis study that the power to detect low frequency asynchrony is low, which undercuts the conclusion that the study has shown that covariance is weak between sardine and anchovy biomass and recruitment (because this result may follow from the lack of power).
The reviewer brings up a tricky issue with our ability to draw conclusions about asynchrony based on limited data. We did find limited power, and do not mean to suggest that asynchrony in landings is strong, only that there is a stronger signal in landings than biomass. In this sense, we also do not want to discount the use of landings for drawing any conclusions from landings data, merely that they should interpret signals of asynchrony cautiously when using only landings data (which are in some ways easier to use for explorations like these, as the datasets for landings are often longer and more complete) We now communicate this in the discussion: "That asynchrony is detectable in some long time series [3] but not others [4] indicates that more data are necessary but not always sufficient for detection, including via our method." (L 596-598).

1.9 For the wavelet analysis methods, the way in which the simulations are constructed is also incompletely described, which obscures the interpretation of results and subsequent conclusions. Nevertheless, the wavelet analysis results are suggestive that landings data, at least at shorter timescales, exhibit stronger negative dependence between sardine and anchovy than either biomass estimates or recruitment data series would suggest. If the wavelet analysis could be better described than this conclusion would be better supported.
We have added a section to the supplement that better describes the wavelet analysis more thoroughly.

1.10 Additionally, the Discussion notes that separating observation and process error might help interpretation of biomass analyses (lines 379-386). This would also suggest that perhaps it's an SSM analysis, which explicitly represents both process and observation error in a parametric model, that could be considered as the primary analysis to base the study conclusions on.
We agree that a state-space model that represents both error types, but struggled to decide how one would interpret the states x_t from a state-space model that has been fitted to estimates that came from a stock assessment. We have also changed the following sentence to clarify that this approach has been previously used to separate observation and process error, by Hosack et al, "A full simulation study that explicitly models observation error could identify the specific roles of process vs. observation error in generating the patterns we found (see [5] as one possible approach)." (L618-620)

Specific comments:
1.11 Lines 105-106: The cited article Izquierdo-Pena et al. (2018) distinguishes 1) synchrony between populations of different regions and 2) alternations between species within regions. In this introductory paragraph, both interpretations (1) and (2) are possible for "asynchrony". The authors should clarify here what is meant by "asynchrony" in their paper: Is it (1) or (2) or both?
We have changed this sentence to say, "A recent study using landings patterns did not find evidence that *sardine-anchovy* asynchrony" (L 108-109)

1.12 Lines 117-119: can remove parentheses here for the cape gannet example
Fixed (this is now L121-123)

1.13 Line 123: " … depends on life history traits and strategies". The distinction between integration over space and time should be made clear. Provide guidance on how integration over time depends on life history traits and strategies in a way that is different from integration over space.
We have removed this sentence

1.14 Line 133-136: run-on sentence
Fixed (now L138-140)

1.15 Lines 148-150: "… surprisingly few attempts …" Although, as suggested by the title, there have been previous attempts to estimate alternation of species with regions and also synchrony of populations among regions. Some of these studies are already cited elsewhere in this paper. One option is to omit "few" here and cite the studies that have looked at these questions previously.
We agree with the reviewer. We do not know of any other studies that have compared different metrics in which one might detect asynchrony (e.g., biomass vs landings vs recruitment) so we have modified the sentence to read, "Despite the substantial attention paid to mechanism, there have been few attempts to characterize this pattern across regions [3,6,7] and to our knowledge none have examined the timescale and prevalence of this pattern across metrics."

1.16 Lines 154-155: "… we analyse the dynamics of the populations themselves …" This sentence is contrasting the current study with the analysis of landings data by Izquierdo-Pena et al. (2018) who argue that landings provide a proxy indicator of abundance or biomass. The suggestion is that the current study has access to alternative data such that the analysis is of the actual populations. The reasoning behind this suggestion should be explained here (later in the paper, the Discussion does note that the biomass time series may not directly correspond to true population abundance).
We have changed this sentence to be more specific, since we do not have raw data or even abundance estimates but use other published time series (biomass and recruitment estimates) to try to get closer to the ecological relationships. The revised sentence reads, "Here, we analyze landings alongside biomass and recruitment as we seek to identify ecological relationships within each ecosystem…" (L165-166)

1.17 Line 184: "… most comprehensive study of trends so far …" This refers again to Izquierdo-Pena et al. (2018), however, there are studies that similarly have looked at historical landings including time series of longer length. It is ambiguous how this citation is the most

comprehensive study. One option is to define what is meant by “most comprehensive”, another is to cite the related studies together.
This sentence has been changed to specify, “This dataset is unique because it examines more variables than the most **globally** comprehensive studies of trends so far [3,6]” – We have specified that we are talking about being geographically comprehensive in terms of studying several ecosystems. We have also cited Hosack et al. here because we believe it fits in the same category of study that seeks a broader pattern.

1.18 Line 190-191: Why define dominant species based on maximum biomass rather than say average or median biomass?
We have changed the comparability analysis to be based on the species with the highest **median** biomass instead of the highest long-term **maximum** (see response to comment 1.1 above).

1.19 Lines 202-203: “… we fit a MARSS model pairs of sardine and anchovy time series …” Clarify that the analysis is applied to sardine and anchovy time series in each region.
The sentence has now been changed to, “To estimate covariance between sardine and anchovy, we fit a MARSS model to the dominant sardine and anchovy stock *in each region*.” (emphasis added here only) (L233-234)

1.20 What are the assumptions of independence among regions?
We assumed that each region has its own variance-covariance matrix for sardine and anchovy (see ESM, equation S1).

1.21 Also, it would be helpful to describe which parameters are assumed identical among regions and why these choices were made.
We originally moved much of the model description to the ESM to save space, but agree that some more detail here is necessary. We describe these in more detail in the Methods (L 291-295).

1.22 Similarly, describe how the different data series (landings, biomass estimates, recruit estimates) are to be analysed, e.g., jointly or separately. How might the three data series provide different perspectives on the same population?
We have added a statement to the Introduction that we are analyzing landings biomass estimates, and recruitment estimates independently: “Here, we analyze landings, biomass and recruitment independently as we seek to identify ecological relationships within each ecosystem”

We describe how the three time series can provide different perspectives on L137-141: “Cycles in landings represent consequences for change in resource use and benefits to human communities, whereas cycles in biomass reflect change in resources available to predators. Finally, recruitment dynamics reflect the underlying production processes that generate biomass and indicates prey availability for predators that specialize on earlier life stages [8,9].”

1.23 Lines 207-209: “ … without phase information.” Insufficient information is provided on how the surrogate simulations were constructed. The citation to the fractal R package is not enough for the simulation procedure to be understood or reproduced.

We now mention the method we used to generate surrogate time series, which is Theiler's phase randomization – it calculates the discrete Fourier transform of the time series and randomizes the phase at each frequency to be uniformly distributed on [0, 2π]. The randomization of the phase information is useful because it provides us with similar time series without relationships to one another, which we use as our "null expectation" for a case where sardine and anchovy biomass, landings, or recruitment are unrelated. We have changed this sentence to read, "Surrogate time series were generated using Theiler's phase randomization using the fractal package in R [10]." Although this does not fully explain the process in the text, a reader interested in replicating the process can search for Theiler's phase randomization and find the function in the fractal package that generates surrogates via this method.

1.24 Lines 222-224: "… somewhat unusual way of looking at wavelet information …" The analysis ignores how values are arranged in the time domain because of the binning by frequency, so why not use Fourier (cross-spectrum) analysis?
Using cross-spectrum analysis is one way we could potentially look at similar information. We chose to use wavelets because this method makes it easier to generate a null distribution with which to compare the observed pattern. We think both of these methods are acceptable ways of looking at the data but wanted to use one that was more interpretable with a statistical test, and the one with which we were more experienced. Because there are several ways to analyze asynchrony in time series, we have added a new section to the supplement called "Why use wavelets?" that introduces some of these methods and explains why we chose to use the method we did for looking at asynchrony. We hope that this will be useful to other researchers pursuing similar questions.
We have also changed the last sentence of this section to say, "In order to compare the observed WMR to a null distribution, we ignored the arrangement of values in the time domain and binned the information by frequency." (L277-278).

1.25 Line 233: The covariance matrix for the process errors should be denoted using a capital letter (because it's a matrix).
We have changed the notation in Eq. 1 to use **Q**, as is traditionally done for state space models (in an earlier draft we needed to distinguish the parameters of the simulation model from those of the model used to estimate covariance, but this is no longer necessary so we have used the notation recommended by the reviewer).

1.26 Line 234--235: Why use "m x 1 vector" and "m x m" matrix when C is a 2 x 1 vector? It must be that m = 2.
We have changed this section to say, "**B** is an $m \times m$ matrix containing the effects of each species on itself along the diagonal and the effect of one species on the other on the off-diagonal (here, $m = 2$). **C** is an $m \times 1$ matrix" – we neglected to write what that m reflects the two species.

1.27 Lines 248-249: "Because we were interested in the detectability … we did not simulate observation error. " If there is an interest in detectability, isn't this an argument in favour of including observation error in the simulations?
We treat the "process error" in our simulations as if it is a combination of process and observation error. This is because a) we don't have a good estimate of observation error to generate "realistic" levels – since our state space model is already being used on a combination

of data and estimated values; b) we are primarily interested in how the time series length influences detectability- with increased observation error we would expect that power would be lower. We have added to L275-276 that "Adding observation error would further decrease the power of the method to detect asynchrony," and mention in the Discussion that our results identify a minimum data requirement for detecting asynchrony (L419).

1.28 Lines 283-288: "Covariance in biomass was similarly weak …" The results focus on the process error covariance matrix but the discussion neglects the non-zero off-diagonal entries of the matrix B that are also related to the cross-correlations among the sardine and anchovy time series within each region and hence likely have joint dependence with the process error covariance matrix estimates. Also Table S4 suggests that some negative dependence could be capture by these off-diagonal entries of B.
It's true that some negative relationship between sardine and anchovy is apparent in the **B** matrix (b_{sa} = -0.10 [-0.19, -0.02]). The reviewer is also correct that the covariance in x_t is a function of **Q** and **B**. We have added material to the supplement that explains how covariance in x_t is produced by these two matrices; the code to replicate the calculations is also included in our GitHub repository (https://github.com/mcsiple/sardine-anchovy), under "QBComparison.R." When we look at the combined effects of **Q** and **B**, they do not result in a net negative covariance between sardine and anchovy in any region where the process error covariances aren't already negative.

1.29 Line 288-230: The range provided for the intraspecific density dependence terms does not include one (equal to density independence). Why is density dependence said to be weak?
We have fixed this—the text now makes it clear that density dependence is weak in sardine (CI includes 1) and present in anchovy (CI does not include 1) (L442-443).

1.30 Lines 303-304: "Across all ecosystems, WMR distributions indicated stronger asynchrony in landings …" Fig. 3b suggests it's unclear that this is true for timescales greater than 10 years in the Japan region.
At the 10+ year scale, it is hard to tell that there is higher density at lower values of WMR in the landings data than for biomass. Here in the text, we now refer the reader to table s3, which gives the confidence intervals for observed – null WMR distributions, with lower values of *d* indicating more asynchrony (if they're negative, it's more than expected by chance). We agree that this is not easy to see in the figure so we have now cited the table as well.

1.31 Lines 313-314: If neither the methods nor the results of an analysis are presented then it is probably better not to report its conclusions (some journals do not allow it). Also remove reference to the analysis in the Discussion. Or add methods/results from the correlation analysis to paper.
These lines in the original text refer to the power analysis described on lines 382-390. While there is limited space in the text to add more description, we have included the methodology into figure 2, to provide a more thorough graphical explanation of methods. We also now report a new table in the supplement (table S5), which shows one of the outcomes of the power analysis. Figures S9 and S15 also show results from the analysis (Figure S15 has been added in this revision). Table S5 and Figure S15 are an attempt to summarize how often the value of d from a Wilcoxon test is within the "true" value for the time series. The power analysis shows two main

things: power to detect asynchrony at very long time scales is low, even with our method (sadly) and the method detects stronger asynchrony than the true level at short time scales. We have tried our best to communicate these results more succinctly in the text.

1.32 Lines 328-332: "… we show that … taxa are not strongly asynchronous … observations of asynchrony in landings are not indicative of underlying patterns in biomass or recruitment." It's unclear that this result was in fact shown. Note that line 308 says that "statistical power to detect asynchrony at decadal scales was low". The reported process covariance terms do not account for the off-diagonal entries of B.
Regarding the off-diagonals of B, we have described in the text now that B has not changed the direction of the estimated covariance (see our response to comment 1.28).

1.33 Lines 337-338: "Significant asynchrony … only occurred in the Northeast Atlantic and Humboldt …" What result(s) is this sentence referring to?
We have clarified that we are referring to the Wilcoxon tests, whose outputs are in table s3 (L 536)

1.34 Lines 342-343: "In the case of the Benguela Current, our finding of synchrony and positive covariance is in agreement with the findings of Izquierdo-Peña et al. [27]." On the contrary, the latter study says that "sardines and anchovies from the Benguela showed alternation in the last part of the last Century" though "In Benguela, alternation between species is not evident during at least the last two decades, as both species have remained at low productivity levels."
This sentence now reads: "In the Benguela Current, our finding of synchrony and positive covariance is in agreement with Izquierdo-Peña et al.'s finding that the stocks began to covary positively in the last two decades after declining to low productivity [6]."

1.35 Lines 383-385: "If observation error is low … it should be possible to improve detection with better data". In general, wouldn't collecting better data reduce observation error?
We have deleted this sentence.

1.36 Eqs. S1 and S2: Why are the marginal process error variances are held constant across the regions whereas the process error correlation is allowed to change?
This was a notation mistake which we are grateful the reviewer caught; separate process error variances are estimated for each region. We have now fixed the equation.

1.37 Why are observation errors held constant across regions? These hypotheses suggest similar levels of environmental stochasticity and observation error for all regions considered, and only correlation in the process error can vary across regions. Also B is held constant across regions. What is the reasoning for these assumptions?
We explain our reasoning behind a shared observation error across regions in our response to reviewer comment 1.3. The environmental stochasticity is characterized by Q matrix for each sardine-anchovy pair, which is allowed to vary across all regions. **B** is held constant for the reasons discussed in our response to comment 1.6; essentially the **B** matrix characterizes interactions within and among species that are separate from environmental effects and other stochasticity, and estimating it allows us to incorporate this variation in the simulations for the power analysis.

1.38 Eq. S5: It’s unclear which parameters vary across regions and which are held fixed.
We have clarified throughout the ESM that for the covariance estimation, the **B** matrix is estimated globally and **Q** is estimated separately for each region; and for the power analysis, the **B** matrix and average values for the elements of **Q** are used in the simulated time series.

1.39 Eq. S6: It’s unclear why the observation error covariance matrix R should depend on B.
This was a notation mistake and has been fixed- the R matrix is diagonal and assumes no observation error covariance between sardine and anchovy (following Eq. S6 in the ESM).

1.40 Last equation and above: equation numbering is out of sequence with references in the captions.
The equation numbering in the ESM has been fixed.

Referee 2

Comments to the Author(s)

2.1 Overall I agree with the authors that previous studies have probably exaggerated the degree of asynchrony in sardines and anchovy given the relatively short time series that are typically used. However the authors could do a better job of justifying their methodological choices. Particularly their definition and analysis of comparability was not well justified.
We have struggled a little with the notion of comparability in terms of contribution to the food web. This measure was intended to be a simple comparison that showed whether one species could hypothetically replace the other one (in terms of total biomass available to predators). Our goal is to show that there is usually one species that dominates biomass and landings. We have revised the description of comparability in the methods to state, “We defined the dominant sardine and dominant anchovy species for each region as the species with the highest median biomass. We assessed **comparability** using the log-ratio of median anchovy to sardine landings and biomass in each ecosystem. This is a simple way of showing whether one species is much more abundant than the other.” (L202-206)

2.2 Title - The methods are not really novel since both the models and data have been used before in different studies. Perhaps a re-analysis of an old story is more accurate.
We have changed the title to simply, “Limited evidence for sardine and anchovy asynchrony: Re-examining an old story”

2.3 Abstract - Both the abstract and introduction should state the number of stocks that were analyzed in this study (five?).
We have added a mention of the five regions in the Abstract (L51), and the Introduction (L169). We also added the total number of time series examined (n = 48) to the Methods section (L198).

2.4 Introduction - A sentence or two should be added to the end of the introduction that mentions the data and models that will be used (MARSS, wavelet, etc.). It seems like the first paragraph of the methods Data section should actually be the last paragraph of the Introduction since it does not really describe the data but rather describes the entire study approach.

We have moved this first paragraph of the Data section up to the beginning of the Methods, since it gives an overview of the approach. We agree with the reviewer that this is a better placement.

Methods-

2.5 Line 191- Why is the log-ratio of maximum landings and biomass a good metric of comparability? Why not choose the mean or median biomass which will be more stable? The maximum seems arbitrary and not particularly ecologically relevant. If this how it has been defined in the past then that should be referenced.

We have changed the metric of comparability to the median of landings and biomass, as recommended by both reviewers. This is reflected in the revised Figure 2 and in the text (L 203-205).

2.6 Line 206 and 252- 50 surrogates seems like a small sample. I would expect hundreds or more. Why just 50? A sensitivity analysis to this would be nice for the supplement.

The simulations are computationally intensive (~8 mins/simulation) and this was the original reason we used few. We repeated the analysis with 100 simulations instead of 50 and have noticed that the result has not changed much (the updated results are based on this instead of the full amount). We have also added another plot to the supplement that shows the result of the full simulation (instead of just a demonstration of one simulation, as figure S9 is)—the new figure S15 shows, for 50 simulations at each observation time series length, the probability that the observed value of asynchrony will fall within the confidence interval of the 'true' value (i.e., the value calculated with the full time series).

2.7 Line 222- If it is unusual then what is usual and why was that not done?

Wavelet information is usually presented with the time domain (as in the heat map in figure S2). In this case, because we wanted to perform statistical tests on the WMR distributions, we had to bin values, essentially collapsing the time domain within each frequency bin. This means that in our statistical comparison, asynchrony at 2 years is not distinguished from asynchrony at 3 or four years, but it is distinguished from asynchrony at a 7- or a 15-year frequency. Figure S2 shows this graphically; we have modified the text in the Methods (L251-252) to simply say: "In order to compare the observed WMR to a null distribution we ignored the arrangement of values in the time domain and binned the information by frequency." It is not unusual enough to merit a special statement to that effect.

2.8 Line 248 - You have estimates of observation error from your MARSS fits to the real data so why not use them in the simulations? Aren't you are interested in the detectability of asynchrony given time series length and realistic observation error? Or perhaps you can state that this is a conservative estimate of time series length since in reality we will have noisy observations which would make the required length even longer.

We originally did not include observation error in the simulations because we wanted to find the lowest possible number of years of data it would take in order to detect asynchrony (i.e., get a conservative estimate). The reviewer is correct about how additional error (process or observation error) would lead to longer times-to-detection.

We have changed the power analysis part of the Methods section to make this clearer, "Adding observation error would further decrease the power of the method to detect asynchrony." (L 285-286) and the Discussion, "In cases where low-frequency asynchrony *is* present, there are significant data requirements for detecting it, of which our estimates represent a minimum." (L423-430).

Results-

2.9 What is the null hypothesis for the comparability analysis? We wouldn't expect the maximums to be exactly equal, so saying they're different doesn't seem very meaningful. If the uncertainty of the ratio could be quantified (perhaps using the mean rather than the maximum) then you could perform a hypothesis test of whether the mean is significantly different from some null.

This is true. Although we didn't use a statistical test to analyze comparability, a null hypothesis would be that the two stocks are of comparable abundance, i.e., it is possible that if their dynamics were compensatory, the biomass of one species could be available to fisheries and predators when the other was not. We now compare the log-ratio of median sardine to anchovy biomass/landings to define which species is dominant, instead of the maximum (L203-204 and updated figure 2).

Discussion -

2.10 Perhaps mention whether you think asynchrony could be detected in the longer geological scale-deposition based time series (e.g., Holmgren-Urba & Baumgartner 1993).

We ran out of time to try this but we have added a digitized version of the scale deposition data to our GitHub repository so that anyone who is curious can see if synchrony is detectable in this dataset using our method or any others.