THE ROYAL SOCIETY
PUBLISHING

PROCEEDINGS B

Landscape complexity promotes resilience of biological pest control to climate change

Benjamin Feit, Nico Blüthgen, Eirini Daouti, Cory Straub, Michael Traugott, and Mattias Jonsson

Article citation details
Proc. R. Soc. B **288**: 20210547.
http://dx.doi.org/10.1098/rspb.2021.0547

Review timeline

Original submission:	9 December 2020
1st revised submission:	5 March 2021
2nd revised submission:	21 April 2021
Final acceptance:	26 April 2021

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

Review History

RSPB-2020-3072.R0 (Original submission)

Review form: Reviewer 1

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

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

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

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

Is the length of the paper justified?
Yes

Should the paper be seen by a specialist statistical reviewer?
Yes

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
Please see attached file. (See Appendix A)

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?
Excellent

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

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

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?
Yes

Is it adequate?
Yes

Do you have any ethical concerns with this paper?
No

Comments to the Author
This manuscript reports a study on resilient services generated by generalist predator communities by a combined analysis on response diversity to temperature gradients and functional redundancy of predator communities. The main conclusion in the abstract is a neat in-a-nutshell summary of the study, i.e. "practices known to strengthen natural pest suppression under current conditions will also confer resilience in ecosystem service provisioning within a changing climate". The study is well written and mainly clearly described, despite the sometimes complex chain of steps taken to derive the analyzed indices and diversity metrics.

I have one concern that I think needs to be addressed before the manuscript can be published. I am wondering if there is a degree of circularity in some of the analyses, most notably between climate resilience to biological control (the response variable) and functional evenness and temperature optima (two response variables). These two interactively explain most variation in climate resilience to biological control, but they are also used to calculate climate resilience to biological control. I was asking myself if this is a problem large enough to cast doubts on the analytical framework, but could not really find an answer. This at least apparent circularity is not discussed in the manuscript, which I think will be needed to justify that the results are valid. Data on climate chambers seem to support this, but I would nevertheless prefer to see a more critical discussion on this aspect. As an example, the in the conclusions this manuscript contributes to showing that trait-based approaches better explain ecosystem functions/services than taxonomic approaches, but I'm wondering how any other result is possible given that the same trait aspects are components in both the response and the descriptors.

On another less major note, the rather complex chain of analytical steps necessitates that terminology is very carefully used, which most often is the case. In particular the variable "resilience to biological control" should be consistently used throughout, it is sometimes referred to in slightly other words, which caused initial confusion for me. Sometimes I lost track on whether response diversity to temperature gradients or climate resilience to biological control was actually referred to. The problem may be that the term resilience itself is in general used in unspecific and incredibly different ways.

Specific comments below.

L42-44 Very neat conclusion, and the potentially important implication from this study (rather than a more general statement about trait-based versus taxonomic approaches).

L69 Avoid starting sentence with the word "this", here it is not clear what you refer to. Do you imply that it is notoriously difficult in farmland and not in other systems? Perhaps restructure sentence.

L73 Ditto. Specify that you refer to a taxonomy-based approach?

L105-107 Rather stress how this study goes beyond Feit et al. 2019 in this sentence, would be more relevant to directly see from the text that these studies have important independent conclusions (if so).

L107-117 Perhaps consider a graphical flow-chart on how different components of this study are used to calculate all the different indices and diversity metrics used? I think this would make it easier to follow the logic of the analysis (and perhaps clarify how trait-space end up in response and explanatory variables).

L115 Add "of biological control" after "climate resilience", if I understood this correctly. I'd strongly recommend to stick to this wording throughout the manuscript when you refer to this variable.

L131-133 Provide a reference for this statement, or, if you got the info from farmers directly, specify this. I do remember that this generally seems to be the case from other contexts.

L141-144 I assume you caught other species of carabids and spiders too. It would be great if you can indicate how high the percentage of the selected species is from the total carabid/spider community.

L146-147 Specify species already here?

L153-155 Specify year?

L176 Super-minor, but I would prefer "species-specific" instead of "predator-specific"

L192 q not specified in the function, assume you refer to p?

L256-264 I found this section hard to follow. Not specified if you analyzed average or card-specific data per trial. The latter would imply that the random structure is wrong (would need trial nested within chamber) so I assume you analyzed average values, but please specify e.g. sample size and response variable. Did you verify normality? In theory values cannot cross the boundaries of 0-1, which could cause problems. Are mean attack rates (L261) based on models on individual cards or means per trial?

L277-278 Entire P(t) function specified on L191 (assume so), or only a part of it (not likely)?

L330-331 Stick to "climate resilience to biological control" for the sake of clarity.

L333 "climate resilience to biological control"

L330-334 This result made me wonder about the potential circularity issue. Functional evenness and climate resilience to biological control (based on functional redundancy) are, if I understand well, calculated from the same underlying trait-based components. I think this potential source of bias should at least be mentioned in the discussion and conclusions, and ideally discussed in a way that would give future research some directions to better cope with finding underlying causalities.

L374 climate resilience to biological control

L380-382 Because aphis consumption is a part of both the response and the predictor in defining a key trait in the system—how could it ever perform worse than taxonomic diversity? Also, would taxonomic diversity beyond the species that you have MGCA-data explain some residual variance? Rarer species could be important for functions despite the fact that they are too few to include in an analysis on whether they consume aphids or not.

L398-403 I agree that there is a prevailing understanding that organic farming is more beneficial in simple rather than in complex landscapes, but there are also studies showing the opposite. In particular I think you could cite Lichtenberg et al. (2017) in this context, at least with an "but see Lichtenberg et al. 2017", but preferably with an explicit sentence notifying that the prevailing idea you refer to would not have been contested. (As an aside, L402: resilience in general, or climate resilience to biological control?)

Lichtenberg E et al. (2017). A global synthesis of diversified farming systems on arthropod diversity within fields and across agricultural landscapes. Global Change Biology 23, 4946-4957.

L438-445 I totally buy the conclusion on L435-438, but these conclusions (L438-445) could be contested (although I agree on a fundamental level!). First, are the results valid and not some spurious effect of a circular argument? Second, you have probably not tested relations between the full array of taxonomic diversity present, but rather the effect of the most dominant species.

Finally. I really enjoyed reading this manuscript, thanks for a very nice study. I hope you find my comments constructive and not overly critical, and that my time-limitations did not cause unintended provocative or rude tone.

Best wishes,
Your reviewer

Decision letter (RSPB-2020-3072.R0)

04-Feb-2021

Dear Dr Feit:

Thank you for the submission of your manuscript RSPB-2020-3072 entitled "Landscape complexity promotes resilience of biological pest control to climate change" to Proceedings B. We have now received referees' reports, and these have been assessed by an Associate Editor.

The manuscript has, in its current form, been rejected for publication. This action has been taken on the advice of referees, who have recommended that substantial revisions are necessary. However we are all agreed that this is a highly interesting topic and that the study has a lot of potential, so we would be happy to consider a resubmission, provided the comments of the referees are fully addressed. However please note that this is not a provisional acceptance.

The resubmission will be treated as a new manuscript. However, we will approach the same reviewers if they are available and it is deemed appropriate to do so by the Editor. Please note that resubmissions must be submitted within six months of the date of this email. In exceptional circumstances, extensions may be possible if agreed with the Editorial Office. Manuscripts submitted after this date will be automatically rejected.

Please find below the comments made by the referees, not including confidential reports to the Editor, which I hope you will find useful. If you do choose to resubmit your manuscript, please upload the following:

1) A ‘response to referees’ document including details of how you have responded to the comments, and the adjustments you have made.
2) A clean copy of the manuscript and one with 'tracked changes' indicating your 'response to referees' comments document.
3) Line numbers in your main document.
4) Data - please see our policies on data sharing to ensure that you are complying (https://royalsociety.org/journals/authors/author-guidelines/#data).

To upload a resubmitted 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 Resubmission." Please be sure to indicate in your cover letter that it is a resubmission, and supply the previous reference number.

Sincerely,
Professor Loeske Kruuk
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
In this study the authors develop and present an index of climate resilience of ecosystem services, using the thermal resilience of predator communities providing biological pest control as an illustration. Both referees had some reservations, which the authors should try to address. Ref 1 points out gaps in the description of the equations, but also queries some of the formulation. Following that, and observed by both referees, they see a circularity of the argument reflected for instance in the equation for climate resilience. In its current form, I'm afraid, I cannot recommend the manuscript for publication.

Reviewer(s)' Comments to Author:
Referee: 1
Comments to the Author(s)
Please see attached file

Referee: 2
Comments to the Author(s)
This manuscript reports a study on resilient services generated by generalist predator communities by a combined analysis on response diversity to temperature gradients and functional redundancy of predator communities. The main conclusion in the abstract is a neat in-a-nutshell summary of the study, i.e. "practices known to strengthen natural pest suppression under current conditions will also confer resilience in ecosystem service provisioning within a changing climate". The study is well written and mainly clearly described, despite the sometimes complex chain of steps taken to derive the analyzed indices and diversity metrics.

I have one concern that I think needs to be addressed before the manuscript can be published. I am wondering if there is a degree of circularity in some of the analyses, most notably between climate resilience to biological control (the response variable) and functional evenness and temperature optima (two response variables). These two interactively explain most variation in climate resilience to biological control, but they are also used to calculate climate resilience to biological control. I was asking myself if this is a problem large enough to cast doubts on the analytical framework, but could not really find an answer. This at least apparent circularity is not discussed in the manuscript, which I think will be needed to justify that the results are valid. Data on climate chambers seem to support this, but I would nevertheless prefer to see a more critical discussion on this aspect. As an example, the in the conclusions this manuscript contributes to showing that trait-based approaches better explain ecosystem functions/services than taxonomic approaches, but I'm wondering how any other result is possible given that the same trait aspects are components in both the response and the descriptors.

On another less major note, the rather complex chain of analytical steps necessitates that terminology is very carefully used, which most often is the case. In particular the variable "resilience to biological control" should be consistently used throughout, it is sometimes referred to in slightly other words, which caused initial confusion for me. Sometimes I lost track on whether response diversity to temperature gradients or climate resilience to biological control was actually referred to. The problem may be that the term resilience itself is in general used in unspecific and incredibly different ways.

Specific comments below.

L42-44 Very neat conclusion, and the potentially important implication from this study (rather than a more general statement about trait-based versus taxonomic approaches).

L69 Avoid starting sentence with the word "this", here it is not clear what you refer to. Do you imply that it is notoriously difficult in farmland and not in other systems? Perhaps restructure sentence.

L73 Ditto. Specify that you refer to a taxonomy-based approach?

L105-107 Rather stress how this study goes beyond Feit et al. 2019 in this sentence, would be more relevant to directly see from the text that these studies have important independent conclusions (if so).

L107-117 Perhaps consider a graphical flow-chart on how different components of this study are used to calculate all the different indices and diversity metrics used? I think this would make it easier to follow the logic of the analysis (and perhaps clarify how trait-space end up in response and explanatory variables).

L115 Add "of biological control" after "climate resilience", if I understood this correctly. I'd strongly recommend to stick to this wording throughout the manuscript when you refer to this variable.

L131-133 Provide a reference for this statement, or, if you got the info from farmers directly, specify this. I do remember that this generally seems to be the case from other contexts.

L141-144 I assume you caught other species of carabids and spiders too. It would be great if you can indicate how high the percentage of the selected species is from the total carabid/spider community.

L146-147 Specify species already here?

L153-155 Specify year?

L176 Super-minor, but I would prefer "species-specific" instead of "predator-specific"

L192 q not specified in the function, assume you refer to p?

L256-264 I found this section hard to follow. Not specified if you analyzed average or card-specific data per trial. The latter would imply that the random structure is wrong (would need trial nested within chamber) so I assume you analyzed average values, but please specify e.g. sample size and response variable. Did you verify normality? In theory values cannot cross the boundaries of 0-1, which could cause problems. Are mean attack rates (L261) based on models on individual cards or means per trial?

L277-278 Entire P(t) function specified on L191 (assume so), or only a part of it (not likely)?

L330-331 Stick to "climate resilience to biological control" for the sake of clarity.

L333 "climate resilience to biological control"

L330-334 This result made me wonder about the potential circularity issue. Functional evenness and climate resilience to biological control (based on functional redundancy) are, if I understand well, calculated from the same underlying trait-based components. I think this potential source of bias should at least be mentioned in the discussion and conclusions, and ideally discussed in a way that would give future research some directions to better cope with finding underlying causalities.

L374 climate resilience to biological control

L380-382 Because aphis consumption is a part of both the response and the predictor in defining a key trait in the system—how could it ever perform worse than taxonomic diversity? Also, would taxonomic diversity beyond the species that you have MGCA-data explain some residual variance? Rarer species could be important for functions despite the fact that they are too few to include in an analysis on whether they consume aphids or not.

L398-403 I agree that there is a prevailing understanding that organic farming is more beneficial in simple rather than in complex landscapes, but there are also studies showing the opposite. In particular I think you could cite Lichtenberg et al. (2017) in this context, at least with an "but see Lichtenberg et al. 2017", but preferably with an explicit sentence notifying that the prevailing idea you refer to would not have been contested. (As an aside, L402: resilience in general, or climate resilience to biological control?)
Lichtenberg E et al. (2017). A global synthesis of diversified farming systems on arthropod diversity within fields and across agricultural landscapes. Global Change Biology 23, 4946-4957.

L438-445 I totally buy the conclusion on L435-438, but these conclusions (L438-445) could be contested (although I agree on a fundamental level!). First, are the results valid and not some spurious effect of a circular argument? Second, you have probably not tested relations between the full array of taxonomic diversity present, but rather the effect of the most dominant species.

Finally. I really enjoyed reading this manuscript, thanks for a very nice study. I hope you find my comments constructive and not overly critical, and that my time-limitations did not cause unintended provocative or rude tone.

Best wishes,
Your reviewer

Author's Response to Decision Letter for (RSPB-2020-3072.R0)

See Appendix B.

RSPB-2021-0547.R0

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?
Good

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

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?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
With this re-submission, the authors have clarified most of my concerns that I had regarding the original submission. However, one of my main concern remains unresolved, which is the only aspect that I will comment on at this stage.

I'm afraid I still cannot buy the validity of the modelling approach relating to attack rates (L 265-270). How is a Poisson distribution suitable to model a response (attack rate) that is expressed on a scale between 0 and 1 (as stated in the response letter), in particular as it is now explicitly stated that average attack rates are modeled? Average values implies that a Poisson response cannot be correct. My original concern was that a normal distribution would not necessarily be fit for a rate (proportion) that varies between 0 and 1. A normal distribution might still be okay given that you would not have a lot of zeroes and ones, but because a Poisson-distribution cannot handle other values than integers, I am even more concerned about the analytical frame than before. If you really model a rate (proportion, or average attacks), a Poisson error must per definition be wrong. Here I think you either need to verify that a Gaussian response is reasonable (as used in the original submission), or if not, use another error distribution (logistic, or beta)—or transform the attack rate (e.g. using arcsin square root) and verify normality.

Review form: Reviewer 3

Recommendation
Accept with minor revision (please list in comments)

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

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

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

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.
No

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
Feit et al. made a very good job with reviewer suggestions. The authors have taken great care to restructuring and clarifying relevant parts of the Methods section, as well as reconsidering some parts of the analysis. I think that the revised manuscript has much improved in clarity and focus. I have only minor comments (these are largely editorial and aimed at clarifying the manuscript – see below) and a potential major suggestion.

I agree with second reviewer that the novel index of climate resilience of biological control is not directly related to measured ecosystem functioning. Although data from feeding trials seem to support your method (the index of thermal resilience developed by the author appears to be related to predator function), I would like to see a more critical discussion about this limitation.

Minor comments

L 74-76 This concept is not very clear. Why might a taxonomic diversity approach hinder the development of conservation strategies?

L 375-380 Although the result is very interesting, not sure if it is so solid for a generalization. In practice, only 10 fields were sampled and testing an interaction with a low number of replications could give unstable model estimates.

L 399-402 This is also another important result that might reflect the insurance hypothesis proposed by Yachi and Loreau (1999). I would spend a few more words on that.

L 417 In my opinion, 'could' would be more appropriate here. Your metric was only tested in 10 fields.

L 419 But in practice, how can this be done? Any example?

Decision letter (RSPB-2021-0547.R0)

12-Apr-2021

Dear Dr Feit,

Thank you for submitting a revised version of this manuscript. This 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 are happy with the majority of the revisions, but there are still some concerns, particularly with regard to the statistical analyses. We would therefore like to invite you to revise your manuscript to address the issues.

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 (https://royalsociety.org/journals/authors/author-guidelines/#data). 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,
Professor Loeske Kruuk
mailto: proceedingsb@royalsociety.org

Associate Editor
Comments to Author:
This is both an easy and a difficult recommendation to make. I agree with the point made by referee one in that for attack rates or even mean attack rates as responses the model applied with Poisson errors seem mis-specified. The authors make no mentioning of having inspected heteroscedasticity, nor do we see mean residual plots, which for the fact that the choice of Poisson errors is very unusual in this situation, would be appropriate. With this model as presented, I can not recommend the manuscript for publication in its current form.
Can the issue be corrected, it certainly can. Is it likely to change the "story", one can not say until the new model results are available, but the differences in figure 3 seem to be reasonably large and they maybe still there even with a different error structure. I do note the support from referee two, but also that this version is already a revision.

Reviewer(s)' Comments to Author:
Referee: 2
Comments to the Author(s).
With this re-submission, the authors have clarified most of my concerns that I had regarding the original submission. However, one of my main concern remains unresolved, which is the only aspect that I will comment on at this stage.

I'm afraid I still cannot buy the validity of the modelling approach relating to attack rates (L 265-270). How is a Poisson distribution suitable to model a response (attack rate) that is expressed on a scale between 0 and 1 (as stated in the response letter), in particular as it is now explicitly stated that average attack rates are modeled? Average values implies that a Poisson response cannot be correct. My original concern was that a normal distribution would not necessarily be fit for a rate (proportion) that varies between 0 and 1. A normal distribution might still be okay given that you would not have a lot of zeroes and ones, but because a Poisson-distribution cannot handle other values than integers, I am even more concerned about the analytical frame than before. If you really model a rate (proportion, or average attacks), a Poisson error must per definition be wrong. Here I think you either need to verify that a Gaussian response is reasonable (as used in the original submission), or if not, use another error distribution (logistic, or beta)—or transform the attack rate (e.g. using arcsin square root) and verify normality.

Referee: 3
Comments to the Author(s).
Feit et al. made a very good job with reviewer suggestions. The authors have taken great care to restructuring and clarifying relevant parts of the Methods section, as well as reconsidering some parts of the analysis. I think that the revised manuscript has much improved in clarity and focus. I have only minor comments (these are largely editorial and aimed at clarifying the manuscript – see below) and a potential major suggestion.

I agree with second reviewer that the novel index of climate resilience of biological control is not directly related to measured ecosystem functioning. Although data from feeding trials seem to support your method (the index of thermal resilience developed by the author appears to be related to predator function), I would like to see a more critical discussion about this limitation.

Minor comments
L 74-76 This concept is not very clear. Why might a taxonomic diversity approach hinder the development of conservation strategies?

L 375-380 Although the result is very interesting, not sure if it is so solid for a generalization. In practice, only 10 fields were sampled and testing an interaction with a low number of replications could give unstable model estimates.

L 399-402 This is also another important result that might reflect the insurance hypothesis proposed by Yachi and Loreau (1999). I would spend a few more words on that.

L 417 In my opinion, ‘could’ would be more appropriate here. Your metric was only tested in 10 fields.

L 419 But in practice, how can this be done? Any example?

Author's Response to Decision Letter for (RSPB-2021-0547.R0)

See Appendix C.

Decision letter (RSPB-2021-0547.R1)

26-Apr-2021

Dear Dr Feit

I am pleased to inform you that your manuscript entitled "Landscape complexity promotes resilience of biological pest control to climate change" 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

Data Accessibility section
Please remember to make any data sets live prior to publication, and update any links as needed when you receive a proof to check. It is good practice to also add data sets to your reference list.

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.

Your article has been estimated as being 9 pages long. Our Production Office will be able to confirm the exact length at proof stage.

Paper charges
An e-mail request for payment of any related charges will be sent out after proof stage (within approximately 2-6 weeks). 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.

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,
Professor Loeske Kruuk
Editor, Proceedings B
mailto: proceedingsb@royalsociety.org

Associate Editor:
Board Member
Comments to Author:
As before the authors have carefully addressed the comments made by the referees and also corrected the statistical model as requested. As we suspected, the overall "story" has not changed. I am pleased to recommend the manuscript for publication in Proc Roy Soc B.

Appendix A

Review of Feit et al.

This manuscript is on an interesting topic attempting to formulate a metric of the resilience of ecosystem function, test its appropriateness and then relate it to landscape structure. The introduction to the paper is clear and well written. However, the combination of studies used, and the methods therein, do not appear to adequately address the aims of the paper. Therefore, although I found the topic and approach very interesting, I have some important reservations about recommending this paper for publication in its current form.

Firstly, a minor point: the second equation is unclear. There is no term '*q*' as described in the text below the equation. Perhaps this is a mistake and should be '*p*'? It is not clear how the probability of predator i feeding on aphids is calculated from the molecular gut content analysis, and this needs explaining. Can you also provide an explanation of the response term P_{ti}, which is missing?

More importantly, the formulation of this equation needs justifying. It seems to be capturing some aspect of the pest control capacity of a given predator at different temperatures, i.e. including probability that aphids are part of the diet, average abundance, body size, etc. The equation includes temperature dependent activity based on 'taxon-specific activation energy' and also the temperature optimum of the species from equation 1. This makes me wonder if there is duplication in this last part of the equation. Doesn't the species temperature optimum implicitly capture temperature dependent activity on your crops, so do you need the 'taxon-specific activation energy' also?

In equation 3 (to calculate redundancy of pest control) it is unclear why you used a Shannon entropy of the pest control capacity across species. Doesn't this give you a measure of the diversity of pest control effectiveness values across all your predators at a given temperature? I would have thought redundancy would relate to some measure of the number of highly effective predators, i.e. on the basis that if one effective predator is absent then others are there to take its functional role; this isn't necessarily captured in a diversity metric? For example, according to your calculation wouldn't three species that are highly effective aphid predators at a given temperature have a lower redundancy score that to highly effective predators plus one very ineffective predator?

Line 262-264 describing calculation of predicted mean attack rates would benefit from an equation as it is a little unclear, in particular this last part 'multiplying measured mean attack rates at one experimental temperature with differences in predation rates between temperatures predicted by our climate niche models"

Results: I can't help feeling the analysis is somewhat circular. For example, you relate diversity in temperature optima to your metric of climate resilience and find a positive relationship (line 327 onwards), yet the equations to calculate climate resilience essentially include temperature optima and a (Shannon) measure of its diversity across predators. In this sense, I'm not sure these results tell us anything unexpected?

Overall, I think the development of a new metric in this paper (cf. first line of discussion: *"We have developed a novel approach to quantify the resilience of ecosystem services to climate change that combines measures of functional redundancy and response diversity into a single metric*") is not adequately tested in terms of its ability to improve upon other measures. It is not related directly to measured ecosystem functioning— the attack rates experiment is related only to the climatic niches of two individual species not this integrated metric.

The metric of resilience is only related to the individual components that make it up in the equation, which is somewhat circular, and not a test of its ability to predict meaningful patterns of ecosystem function in the real world. This means that strong statements in the discussion such as "*These findings provide further evidence that trait-based approaches are better suited than taxonomic diversity to predict biodiversity effects on ecosystem services*" seem poorly justified.

Appendix B

Board Member: 1
Comments to Author:
In this study the authors develop and present an index of climate resilience of ecosystem services, using the thermal resilience of predator communities providing biological pest control as an illustration. Both referees had some reservations, which the authors should try to address. Ref 1 points out gaps in the description of the equations, but also queries some of the formulation. Following that, and observed by both referees, they see a circularity of the argument reflected for instance in the equation for climate resilience. In its current form, I'm afraid, I cannot recommend the manuscript for publication.

Authors:
Thank you for the invitation to revise and resubmit our manuscript for further evaluation. We have now addressed all comments raised during the review process. We highly appreciate the effort by both Referees and yourself and believe that the concerns raised during the review process have helped us to greatly improve clarity, methodology and focus of our work.

In particular, we agree with the concerns of both referees that the circularity of the community driver analysis was a major weakness of the manuscript. As the main focus of our work – the development of a new approach to calculate resilience of ecosystem services exemplified by climate resilience in biological control – does not rely on this part of the analysis, we have now removed this section.

We have also addressed the concerns regarding the calculation of temperature-dependent risk of predation by removing the temperature-dependency of the metabolic rate, as suggested by Referee 2. As a consequence, the results of the analysis changed only marginally, further confirming that including temperature niches for in-field activity is sufficient to model temperature-dependent aphid consumption probabilities by individual species. We have thus now amended the methodology as well as relevant figures and sections of the results.

Further, we have restructured relevant parts of the Methods section to improve clarity in regards to sampling procedures, climate resilience calculations and mesocosm experiments and now provide a flowchart illustrating the links between species-specific traits, redundancy, and climate resilience of biological control services (Fig. 1).

Please see our detailed responses to each of the points raised by the Referees below.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
Please see attached file

Referee: 2

Comments to the Author(s)
This manuscript reports a study on resilient services generated by generalist predator communities by a combined analysis on response diversity to temperature gradients and functional redundancy of predator communities. The main conclusion in the abstract is a neat in-a-nutshell summary of the study, i.e. “practices known to strengthen natural pest suppression under current conditions will also

confer resilience in ecosystem service provisioning within a changing climate”. The study is well written and mainly clearly described, despite the sometimes complex chain of steps taken to derive the analyzed indices and diversity metrics.

Authors:
Thank you for this positive assessment.
Please note that, in addition to the changes made in response to the specific comments below, we also restructured the field sampling component in the Methods section to improve clarity in regards to sampling procedures (L121-166).

I have one concern that I think needs to be addressed before the manuscript can be published. I am wondering if there is a degree of circularity in some of the analyses, most notably between climate resilience to biological control (the response variable) and functional evenness and temperature optima (two response variables). These two interactively explain most variation in climate resilience to biological control, but they are also used to calculate climate resilience to biological control. I was asking myself if this is a problem large enough to cast doubts on the analytical framework, but could not really find an answer. This at least apparent circularity is not discussed in the manuscript, which I think will be needed to justify that the results are valid. Data on climate chambers seem to support this, but I would nevertheless prefer to see a more critical discussion on this aspect. As an example, the in the conclusions this manuscript contributes to showing that trait-based approaches better explain ecosystem functions/services than taxonomic approaches, but I’m wondering how any other result is possible given that the same trait aspects are components in both the response and the descriptors.

Authors:
The question we tried to answer with the analysis of community determinants of climate resilience was how do the different community determinants the metric is built upon (i.e., taxonomic diversity, functional diversity, and response diversity) contribute to resilience? The circularity in the analysis was thus by design as we explicitly investigated the relative importance of each component the metric consists of.
However, both reviewers have criticized this approach and found it to be a major weakness of the manuscript. As the main focus of our work – the development of a new approach to calculate resilience of ecosystem services exemplified by climate resilience in biological control – does not rely on this part of the analysis, we have now removed this section from the manuscript. Numerous previous studies have already demonstrated the importance of trait-based approaches and the motivation to use indicators of trait diversity over simple taxonomic diversity metrics is sufficiently provided in the introduction and discussion sections.

On another less major note, the rather complex chain of analytical steps necessitates that terminology is very carefully used, which most often is the case. In particular the variable “resilience to biological control” should be consistently used throughout, it is sometimes referred to in slightly other words, which caused initial confusion for me. Sometimes I lost track on whether response diversity to temperature gradients or climate resilience to biological control was actually referred to. The problem may be that the term resilience itself is in general used in unspecific and incredibly different ways.

Authors:
Thank you for pointing this out – we have now amended the manuscript and use the term “resilience of biological control” throughout to improve clarity.

Specific comments below.

L42-44 Very neat conclusion, and the potentially important implication from this study (rather than a more general statement about trait-based versus taxonomic approaches).

Authors:
Thank you for this compliment.

L69 Avoid starting sentence with the word “this”, here it is not clear what you refer to. Do you imply that it is notoriously difficult in farmland and not in other systems? Perhaps restructure sentence.

Authors:
Now L70. We have amended relevant sections in the paragraph and now avoid “this” at the beginning of a sentence.

L73 Ditto. Specify that you refer to a taxonomy-based approach?

Authors:
Now L74. We have restructured the sentence accordingly and now refer to taxonomy-based approaches directly.

L105-107 Rather stress how this study goes beyond Feit et al. 2019 in this sentence, would be more relevant to directly see from the text that these studies have important independent conclusions (if so).

Authors:
L105-108. We have amended the section to now emphasize the difference between the metric in Feit et al. 2019, which allows for the calculation of redundancy in current predator communities, and the present study, which allows predictions of resilience to future changes in climatic conditions: “Here, we extend a recently introduced metric of functional redundancy in biological control services under current conditions [6] by including a measure of response diversity to ambient temperatures in order to calculate the resilience of biological control services to increased temperature variability under future climate conditions.”

L107-117 Perhaps consider a graphical flow-chart on how different components of this study are used to calculate all the different indices and diversity metrics used? I think this would make it easier to follow the logic of the analysis (and perhaps clarify how trait-space end up in response and explanatory variables).

Authors:
Thank you very much for this excellent suggestion. We have now added a flow chart (Fig. 1) and believe that it greatly improves the interpretability of our resilience metric.

L115 Add “of biological control” after “climate resilience”, if I understood this correctly. I’d strongly recommend to stick to this wording throughout the manuscript when you refer to this variable.

Authors:
L115. As suggested, we have amended the wording here and at appropriate places throughout the manuscript to improve clarity.

L131-133 Provide a reference for this statement, or, if you got the info from farmers directly, specify this. I do remember that this generally seems to be the case from other contexts.

Authors:
L131-133. Thank you for pointing this out. We now cite Weibull et al 2003 as a reference for this statement.

L141-144 I assume you caught other species of carabids and spiders too. It would be great if you can indicate how high the percentage of the selected species is from the total carabid/spider community.

Authors:
L190-194. We have now added the suggested data: "Analyses of the climate resilience of biological control were conducted on a subset of arthropod predator communities and comprised seven wolf spider species (1,775 individuals, 86.9% of all captures) and nine carabid beetle species (4,722 individuals, 92.2% of all captures) for which we were able to gather information on all three field-monitored components of the resilience metric: abundance, gut content and temperature niche."

L146-147 Specify species already here?

Authors:
Please see our response to the comment above.

L153-155 Specify year?

Authors:
L150. We have now added the year of sampling.

L176 Super-minor, but I would prefer "species-specific" instead of "predator-specific"

Authors:
We have changed the wording accordingly throughout the manuscript.

L192 q not specified in the function, assume you refer to p?

Authors:
Now L202. This is correct. We changed the letter accordingly.

L256-264 I found this section hard to follow. Not specified if you analyzed average or card-specific data per trial. The latter would imply that the random structure is wrong (would need trial nested within chamber) so I assume you analyzed average values, but please specify e.g. sample size and response variable. Did you verify normality? In theory values cannot cross the boundaries of 0-1, which could cause problems. Are mean attack rates (L261) based on models on individual cards or means per trial?

Authors:
We used average attack rates for each container during each temperature treatment. This information is now added to the text (L265-266).
L273-280. Further, we added the following section to the manuscript to improve clarity: "Predicted mean attack rates (*PM*) were calculated for each design by multiplying measured mean attack rates at the optimal experimental temperature with expected differences in predation rates between temperatures as predicted by the approximated risk of predation (*P*) at both temperatures:

$$PM_{T_{subopt},i} = M_{T_{opt},i} \cdot (\frac{P_{T_{opt},i}}{P_{T_{subopt},i}})$$

Where *M* is the measured attack rate of species i at its temperature optimum T_{opt} and *P* the approximated risk of predation by species i at their respective temperature optima T_{opt} and sub-optima T_{subopt}."
Additionally, we are now using a Poisson distribution in our modelling approach to address the problem of values not being able to cross 0-1. The new analysis only marginally changed the F values and the overall results remained the same.

L277-278 Entire P(t) function specified on L191 (assume so), or only a part of it (not likely)?

Authors:
This part of the analysis has now been removed.

L330-331 Stick to "climate resilience to biological control" for the sake of clarity.
L333 "climate resilience to biological control"
L374 climate resilience to biological control

Authors:
Thank you for pointing these out. As stated above, we have changed the wording throughout the manuscript to improve clarity.

L330-334 This result made me wonder about the potential circularity issue. Functional evenness and climate resilience to biological control (based on functional redundancy) are, if I understand well, calculated from the same underlying trait-based components. I think this potential source of bias should at least be mentioned in the discussion and conclusions, and ideally discussed in a way that would give future research some directions to better cope with finding underlying causalities.
L380-382 Because aphis consumption is a part of both the response and the predictor in defining a key trait in the system—how could it ever perform worse than taxonomic diversity? Also, would taxonomic diversity beyond the species that you have MGCA-data explain some residual variance? Rarer species could be important for functions despite the fact that they are too few to include in an analysis on whether they consume aphids or not.
L438-445 I totally buy the conclusion on L435-438, but these conclusions (L438-445) could be contested (although I agree on a fundamental level!). First, are the results valid and not some spurious effect of a circular argument? Second, you have probably not tested relations between the full array of taxonomic diversity present, but rather the effect of the most dominant species.

Authors:
These sections of the analysis and discussion have now been removed from the manuscript. Please see specific comment above.

L398-403 I agree that there is a prevailing understanding that organic farming is more beneficial in simple rather than in complex landscapes, but there are also studies showing the opposite. In particular I think you could cite Lichtenberg et al. (2017) in this context, at least with an "but see Lichtenberg et al. 2017", but preferably with an explicit sentence notifying that the prevailing idea you refer to would not have been contested. (As an aside, L402: resilience in general, or climate resilience to biological control?)
Lichtenberg E et al. (2017). A global synthesis of diversified farming systems on arthropod diversity within fields and across agricultural landscapes. Global Change Biology 23, 4946-4957.

Authors:
L377-378. We have now added the recommended reference to provide the reader with the information that the often reported stronger effects of organic farming on taxonomic diversity metrics in simple rather than complex landscapes has also been contested: "…which is in contrast to frequently reported [40,41], but not uncontested [42], stronger increases in taxonomic diversity in organic farms located in simplified landscapes"
L379-380. In regards to the side note: Resilience is here referred to as for ecosystem services in general, we have now clarified this in the text.

Finally. I really enjoyed reading this manuscript, thanks for a very nice study. I hope you find my comments constructive and not overly critical, and that my time-limitations did not cause unintended provocative or rude tone.
Authors:
Thank you very much for your detailed evaluation of our manuscript. We believe that your comments helped us to significantly improve the interpretability and quality of our work.

Best wishes,

Your reviewer

Referee 2

Review of Feit et al.
This manuscript is on an interesting topic attempting to formulate a metric of the resilience of ecosystem function, test its appropriateness and then relate it to landscape structure. The introduction to the paper is clear and well written. However, the combination of studies used, and the methods therein, do not appear to adequately address the aims of the paper. Therefore, although I found the topic and approach very interesting, I have some important reservations about recommending this paper for publication in its current form.

Thank you for your thoughtful comments and suggestions that helped us to strengthen both the analytical approach and focus of our manuscript.
Please note that, in addition to the changes made in response to the specific comments below, we also restructured the field sampling component in the Methods section to improve clarity in regards to sampling procedures (L121-166).

Firstly, a minor point: the second equation is unclear. There is no term 'q' as described in the text below the equation. Perhaps this is a mistake and should be 'p'? It is not clear how the probability of predator i feeding on aphids is calculated from the molecular gut content analysis, and this needs explaining. Can you also provide an explanation of the response term Pti, which is missing?
More importantly, the formulation of this equation needs justifying. It seems to be capturing some aspect of the pest control capacity of a given predator at different temperatures, i.e. including probability that aphids are part of the diet, average abundance, body size, etc. The equation includes temperature dependent activity based on 'taxon-specific activation energy' and also the temperature optimum of the species from equation 1. This makes me wonder if there is duplication in this last part of the equation. Doesn't the species temperature optimum implicitly capture temperature dependent activity on your crops, so do you need the 'taxon-specific activation energy' also?

Authors:
L202. The 'q' was indeed an error and should have read 'p'. Thank you for pointing this out, we have now corrected to 'p'.

L202-203. The probability was calculated as the frequency of detection of aphid DNA in the gut of the specimens of a respective species. The MGCA allowed for simple presence/absence data for prey DNA and the probability of predation is calculated as positives divided by total specimens. We have now added this information to the text "expressed as the fraction of specimens within a species tested positive for aphid DNA in their gut content"

L198. We have also now added the missing explanation of the response term $P_{T,i}$.

We agree with the point made that the temperature niche captured by field surveys already encompasses temperature-dependent activity. Consequently, the temperature-related scaling term for the metabolic rate calculations might not be necessary. After removing the term for temperature-dependency of the metabolic rate, the results of the analysis only changed marginally, further confirming that including temperature niches for in-field activity is sufficient to model temperature-dependent aphid consumption probabilities by individual species. We have thus now amended the methodology as well as relevant figures and sections of the results.

In equation 3 (to calculate redundancy of pest control) it is unclear why you used a Shannon entropy of the pest control capacity across species. Doesn't this give you a measure of the diversity of pest control effectiveness values across all your predators at a given temperature? I would have thought redundancy would relate to some measure of the number of highly effective predators, i.e. on the basis that if one effective predator is absent then others are there to take its functional role; this isn't necessarily captured in a diversity metric? For example, according to your calculation wouldn't three species that are highly effective aphid predators at a given temperature have a lower redundancy score that to highly effective predators plus one very ineffective predator?

Authors:
It is correct that our approach to calculate functional redundancy results in a measure of the diversity of pest control effectiveness (expressed as risk of predation $P_{T,i}$ for aphids to individual predator species i) across all predators at a given temperature. However, the index we use, Shannon entropy, represents, in its original form, a measure of how much certainty there is when predicting the species identity of a randomly chosen individual given its abundance in a dataset. As a result, a higher Shannon means more equal representation of species in a community while lower equates to less equal representation. Adapting this metric to calculate functional redundancy based on the indices of risk of predation in a community, a higher value represents equal contribution of species to the service at a given temperature, while a lower value means unequal contribution. Using the exponential

Shannon, which follows a linear distribution and has a doubling property, then allows for a direct comparison of redundancy between communities. Its value approaches 1 in dissimilar communities and equals N (i.e., the total number of species) in communities consisting entirely of species that are identical in their niche dimensions (i.e., same temperature niche, metabolic rate, aphid frequency in guts and abundance). As a result, a community consisting of three predators with very similar niche dimensions would approach a functional redundancy value of 3, whereas the functional redundancy of a community consisting of three predators with two with very similar niche dimensions and one with dissimilar niche dimensions would be between 2 and 3 but lower than the first example. Thus, our functional redundancy metric adequately assigns higher values to communities characterized by high functional niche overlap and lower values to communities characterized by low functional niche overlap at a given temperature.

Line 262-264 describing calculation of predicted mean attack rates would benefit from an equation as it is a little unclear, in particular this last part ‘multiplying measured mean attack rates at one experimental temperature with differences in predation rates between temperatures predicted by our climate niche models”

Authors:
L273-280. Thank you for suggesting this improvement. We have now clarified the calculation by providing the following equation:
“Predicted mean attack rates (*PM*) were calculated for each design by multiplying measured mean attack rates at the optimal experimental temperature with expected differences in predation rates between temperatures as predicted by the approximated risk of predation (*P*) at both temperatures:

$$PM_{T_{subopt},i} = M_{T_{opt},i} \cdot (\frac{P_{T_{opt},i}}{P_{T_{subopt},i}})$$

Where *M* is the measured attack rate of species i at its temperature optimum T_{opt} and *P* the approximated risk of predation by species i at their respective temperature optima T_{opt} and sub-optima T_{subopt}.”

Results: I can’t help feeling the analysis is somewhat circular. For example, you relate diversity in temperature optima to your metric of climate resilience and find a positive relationship (line 327 onwards), yet the equations to calculate climate resilience essentially include temperature optima and a (Shannon) measure of its diversity across predators. In this sense, I’m not sure these results tell us anything unexpected?

Authors:
The question we tried to answer with the analysis of community determinants of climate resilience was how do the different community determinants the metric is built upon (i.e., taxonomic diversity, functional diversity, and response diversity) contribute to resilience? The circularity in the analysis was thus by design as we explicitly investigated the relative importance of each component the metric consists of.
However, both referees have criticized this approach and found it to be a major weakness of the manuscript. As the main focus of our work – the development of a new approach to calculate resilience of ecosystem services exemplified by climate resilience in biological control – does not rely on this part of the analysis, we have now removed this section from the manuscript. Numerous previous studies have already demonstrated the importance of trait-based approaches and the motivation to use indicators of trait diversity over simple taxonomic diversity metrics is sufficiently provided in the introduction and discussion sections.

Overall, I think the development of a new metric in this paper (cf. first line of discussion: *“We have* *developed a novel approach to quantify the resilience of ecosystem services to climate change that combines measures of functional redundancy and response diversity into a single metric*") is not adequately tested in terms of its ability to improve upon other measures. It is not related directly to measured ecosystem functioning— the attack rates experiment is related only to the climatic niches of two individual species not this integrated metric.

Authors:
The analysis of the attack rate experiment consists of an analysis of differences in predation rates by both species at their respective optimum and suboptimum temperatures and a comparison between the measured rates and the rates predicted by our formula for the risk of predation $P_{T,i}$. The calculation of the latter is based on the species-specific aphid frequency, metabolic rate, temperature niche and abundance (here, as a constant because abundances did not differ between treatments). It is thus a direct test of the predictive power of a fundamental component of the resilience metric presented here. The predictions generated by the calculation are in line with the measured differences in attack rates between temperature treatments and thus confirm that the metric is built upon service-relevant traits and suitable to calculate species-specific differences in service provision. How the predictions were generated, however, was not clear in the previous version of the manuscript, as also pointed out by Reviewer 1. This has now been improved to show the link between the attack rate experiments and the conceptual framework of our metric of resilience more clearly (L273-280, see response to the specific comment above).

We agree that we did not exhaustively test the predictive power of our resilience metric and only demonstrated the link between the conceptual framework and actual pest control potential at the example of two carabids in our system. We have now acknowledged this limitation (L392-397): "A limitation of this study is that we validated the climate niche models with only two species. However, all the other predators in our study use an active hunting strategy similar to P. cupreus and P. melanarius and show similarly distinct climate niches, so it is reasonable to expect that their temperature-dependent consumption patterns can be generalized to other predators in our study system."

Numerous studies have demonstrated that trait-based approaches are better suited to investigate ecosystem functioning than purely taxonomic metrics and that there is a major deficit in our current understanding of resilience drivers and a lack of approaches to quantify resilience. The confirmation of the conceptual framework by our attack rate experiment demonstrates a clear link between theoretical resilience and pest control. Consequently, we believe that our calculation of climate resilience as a function of taxonomy and trait-based functional redundancy over a temperature gradient constitutes a clear advancement over many previous approaches.

The metric of resilience is only related to the individual components that make it up in the equation, which is somewhat circular, and not a test of its ability to predict meaningful patterns of ecosystem function in the real world. This means that strong statements in the discussion such as "*These findings provide further evidence that trait-based approaches are better suited than taxonomic diversity to predict biodiversity effects on ecosystem services*" seem poorly justified.

Authors:
As both reviewers found our circular approach to calculate the relative importance of community determinants of resilience to be a major weakness of the manuscript, we have removed the analysis and relevant sections (such as the above statement) from the manuscript.

Appendix C

Associate Editor
Comments to Author:
This is both an easy and a difficult recommendation to make. I agree with the point made by referee one in that for attack rates or even mean attack rates as responses the model applied with Poisson errors seem mis-specified. The authors make no mentioning of having inspected heteroscedasticity, nor do we see mean residual plots, which for the fact that the choice of Poisson errors is very unusual in this situation, would be appropriate. With this model as presented, I can not recommend the manuscript for publication in its current form.
Can the issue be corrected, it certainly can. Is it likely to change the "story", one can not say until the new model results are available, but the differences in figure 3 seem to be reasonably large and they maybe still there even with a different error structure. I do note the support from referee two, but also that this version is already a revision.

Authors:
Thank you very much for the invitation to revise and resubmit our manuscript. We completely agree with the assessment of Referee 2 and yourself, that the modelling approach for the attack rate experiment was inappropriate. However, after unnecessary confusion, this was an easy fix as the experimental design was specifically intended to allow for a simple and straightforward analysis using GLMM with a binomial distribution. Each design (two single species and one multi species) was replicated in 15 containers under cold and 15 containers under warm conditions. Over the course of 5 h, each replicate was presented with 10 aphid cards at a rate of 1 card every 30 minutes. After each 30-minute interval, the previous card was replaced by a new card and checked for predation. An attack was defined as, at minimum, partial consumption of the aphids presented on a card. This means that each measurement had a binary outcome, i.e., 1 for predation, 0 for no predation.

We have fixed the analysis as follows:
We now use a GLMM with a binomial distribution to investigate differences in the attacks on aphid cards between temperature treatments in each design. Container ID was entered as a random factor to account for replication within containers.

We have made amendments to the methods (L269-283) and results sections (L328-335) accordingly. As the relative difference between treatments remained the same the new analysis did not require changes to the discussion.

We have further addressed the suggestions by Referee 3 and made relevant changes to the manuscript. Please see our response to each point raised by Referee 3 below.

Reviewer(s)' Comments to Author:

Referee: 2

Comments to the Author(s).
With this re-submission, the authors have clarified most of my concerns that I had regarding the original submission. However, one of my main concern remains unresolved, which is the only aspect that I will comment on at this stage.

I’m afraid I still cannot buy the validity of the modelling approach relating to attack rates (L 265-270). How is a Poisson distribution suitable to model a response (attack rate) that is expressed on a scale between 0 and 1 (as stated in the response letter), in particular as it is now explicitly stated that average attack rates are modeled? Average values implies that a Poisson response cannot be correct. My original concern was that a normal distribution would not necessarily be fit for a rate (proportion) that varies between 0 and 1. A normal distribution might still be okay given that you would not have a lot of zeroes and ones, but because a Poisson-distribution cannot handle other values than integers, I am even more concerned about the analytical frame than before. If you really model a rate (proportion, or average attacks), a Poisson error must per definition be wrong. Here I think you either need to verify that a Gaussian response is reasonable (as used in the original submission), or if not, use another error distribution (logistic, or beta)—or transform the attack rate (e.g. using arcsin square root) and verify normality.

Authors:
We entirely agree that the modelling approach was wrong and have now amended this. Please see our comment above.

Referee: 3

Comments to the Author(s).
Feit et al. made a very good job with reviewer suggestions. The authors have taken great care to restructuring and clarifying relevant parts of the Methods section, as well as reconsidering some parts of the analysis. I think that the revised manuscript has much improved in clarity and focus. I have only minor comments (these are largely editorial and aimed at clarifying the manuscript – see below) and a potential major suggestion.

Authors:
Thank you very much for this positive assessment of our revision. And also thank you for raising the following points that have helped us to further improve the clarity of our manuscript.

I agree with second reviewer that the novel index of climate resilience of biological control is not directly related to measured ecosystem functioning. Although data from feeding trials seem to support your method (the index of thermal resilience developed by the author appears to be related to predator function), I would like to see a more critical discussion about this limitation.

Authors:
We have now added an additional caveat to the discussion:
L397-399 “We therefore caution that additional experiments are warranted to further corroborate the postulated link between response diversity and resilience of biological control services to future climate conditions.”

Minor comments

L 74-76 This concept is not very clear. Why might a taxonomic diversity approach hinder the development of conservation strategies?

Authors:
Thank you for pointing this out. We have now added further clarification:
L 74-77: "Because the trait and response diversity of ecosystem service providers are directly related to how they function, and taxonomic diversity is only indirectly related, a focus on the former may better enable ecologists to engineer ecosystems for resilience through carefully targeted conservation strategies [2]"

L 375-380 Although the result is very interesting, not sure if it is so solid for a generalization. In practice, only 10 fields were sampled and testing an interaction with a low number of replications could give unstable model estimates.

Authors:
In this section we note that we found positive effects of organic farming on climate resilience to be strongest in complex landscapes. These findings contrast frequent (albeit not uncontested) reports of effects of organic farming to be most beneficial for taxonomic diversity in simplified landscapes. We then state that this discrepancy shows that using species diversity as a proxy for resilience could lead to erroneous conclusions and provide two references further backing this claim [43,44]. Together with previous evidence on the problems of using taxonomic diversity as a proxy for resilience, our study provides further evidence and additional insight into why this is likely the case. We agree that the limited number of replicates could result in unstable model estimates and therefore limited generalisability of results. We have now toned down the wording to put our results more in relation to previous findings:
L381-383 "This discrepancy provides further evidence to the notion that measures of trait and response diversity may better inform conservation strategies targeting the preservation of ecological resilience than taxonomic diversity [2,43-46]."

L 399-402 This is also another important result that might reflect the insurance hypothesis proposed by Yachi and Loreau (1999). I would spend a few more words on that.

Authors:
Thank you for this suggestion. We have now added the reference and the following sentence to the discussion:
L408-410 "These findings are consistent with the insurance hypothesis which states that systems characterised by greater response diversity are more resilient to disturbance [47]."

L 417 In my opinion, 'could' would be more appropriate here. Your metric was only tested in 10 fields.
Authors:

L425 We agree and have changed the wording to “could”.

L 419 But in practice, how can this be done? Any example?

Authors:
We have added an additional sentence detailing management strategies:
L428-431 “This might be achieved by increasing the availability and diversity of permanent habitat and temporary refugia to bolster a diversity of seemingly redundant species already present in the system but could also require assisted migration and/or targeted reintroduction of species [3,49,50].”