THE ROYAL SOCIETY PUBLISHING

PROCEEDINGS B

Crop production in the USA is frequently limited by a lack of pollinators

J. R. Reilly, D. R. Artz, D. Biddinger, K. Bobiwash, N. K. Boyle, C. Brittain, J. Brokaw, J. W. Campbell, J. Daniels, E. Elle, J. D. Ellis, S. J. Fleischer, J. Gibbs, R. L. Gillespie, K. B. Gundersen, L. Gut, G. Hoffman, N. Joshi, O. Lundin, K. Mason, C. M. McGrady, S. S. Peterson, T. L. Pitts-Singer, S. Rao, N. Rothwell, L. Rowe, K. L. Ward, N. M. Williams, J. K. Wilson, R. Isaacs and R. Winfree

Article citation details
Proc. R. Soc. B **287**: 20200922.
http://dx.doi.org/10.1098/rspb.2020.0922

Review timeline

Original submission:	11 October 2019
1st revised submission:	23 April 2020
2nd revised submission:	25 June 2020
Final acceptance:	7 July 2020

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

Review History

RSPB-2019-2393.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?
Poor

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

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

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?
Yes

Comments to the Author
I realized when I started to review this submission that it was a version of a paper that I previously reviewed at another journal (PNAS) and where I raised quite a number of concerns. Now that I examine the ms for review I realise that many of my concerns relate not just to this version, but to changes that have been made (and many that have not been made) since the last version that I saw. In particular I have picked up a change that requires some explanation.

The original ms had the following text regarding methods for assessing apple fruit set "Although apples are typically thinned to achieve fruit that meet fresh-market standards, all apple pollination assessments were **before** this to avoid the variation among farms in their thinning practices." (emphasis added) This prompted me to raise the following concern in my original review
"Getting a true yield measure is difficult ……..This problem is greatest for apples, where on line 206 it is explained that fruit set was assessed before the crop was thinned. Assessing fruit set this way is certainly more likely to detect pollen limitation than a post thinning assessment, but the response variable is not commercial yield. It is quite possible that, for example there is a significant relationship between flower visits and early fruit set (such as was found, line 788) but no relationship with yield of mature fruit."
This revised ms has text in the methods that indicates a change since the PNAS version, perhaps in response to the concern that I raised. Here is the text:
"Apples are typically thinned to achieve fruit that meet fresh-market standards; thus our apple fruit counts were taken **post-thinning** to be more directly related to harvestable yield. This is a conservative approach, because post-thinning measurements are less likely than those taken pre-thinning to detect the effect of pollination limitation." (emphasis added).
I was surprised to see this, because it suggested that both pre-thinning and post-thinning assessment were done, and therefore the authors were able to re-analyse using a revised data set. Then I examined the figures (fig s2) and noted that the data were exactly the same as that presented in the earlier version, even though the methods indicated a change in approach. I can think of at three possible explanations for this:
1) It may be that the first version I read was in fact wrong with regard the method and therefore it is possible that this revised version is correct, whereas the earlier version was wrong.
2) It may be that the authors made a mistake in submitting the old figures in the revised ms.

3) It may be that the authors changed the method text but failed to run a new analysis.

Other concerns raised in my original review, and which remain a concern

Table S3 reveals that the sampling effort for flower observations is quite small in many cases. In a number of crops observations were only made on one date per year and at only one time of day. For a number of crops the total time spent observing in each transect was only 6.7 minutes. Given that pollinators are active for many hours of each day, and flowering for most crops extends over more than a week, one would have to expect that this level of sampling will have limited power to detect patterns. Adding more noise, it is known that not all visits are equal (in terms of how the flower is handled by bees, for example).

Because of the difficulty in getting comprehensive observations on visitor frequency over the life of flowers or the full flowering season it is not unexpected that pollen limitation might be detected by other methods (like distance from hive effects and pollen supplementation effects) in that same sites that flower visitor observations fail to detect a pattern. Other studies have shown this. It is certainly apparent that even where positive relationships are detected in this study, they are generally not tight fits (ie large y residuals in the graphs of fig s2).

This problem is likely to lead to conservative conclusions, e.g. they are likely to fail to detect pollen limitation even when it is present. It is also expected that the relatively poor fit of the relationships also makes estimates of the "breakpoint" (i.e. the threshold in relationships that first increase and then flatten) is likely to have wide error margins. The analytical methods do not explore the consequences of this problem.

In the paragraph beginning line 142 (and then throughout) the paper refers to the idea that the study assesses pollen limitation, when in fact the method measures visitor limitation. There is an unstated assumption that the two are almost the same, but in agricultural systems we know that pollen limitation can also be a consequence of planting design especially in crops that have self-infertile varieties which require cross pollination between different rows.

Other studies have explicitly examined the importance of pollinator diversity in reducing likelihood of pollen limitation in two quite different ways, i.e. spatially and temporally. This study does not to examine the temporal component (no exploration of time of day, time of season or year effects). The variability examined is therefore only spatial (transect to transect and site to site). The text should be more explicit about this design decision and the consequences.

The paper uses the term "yield" throughout, which is generally defined as the mass of the crop produced per unit area of farmed land. It is a powerful variable because it relates very closely to efficiency of production. Because one of the aims of the study is to understand the economic value of crop pollination, it is particularly important to use measures of production that relate closely to commercial outcomes. Across the different production metrics in this study only two crops (Table s4, pumpkin and melon) were assessed on a per area basis. For most crops in the study the measures were fruit per flower or branch, which are well short of true yield. Getting a true yield measure is difficult in the experimental context, so this is not surprising. Nevertheless it is important to be careful in the choice of words.

Other studies show that there can be large differences in breeding system (and therefore vulnerability to pollen limitation) among cultivars or varieties of the same crop. However, the method description is not clear regarding how differences in cultivar are handled, except the comment on line 29 that implies the "regionally dominant" cultivar was examined. The description of the model fitting in the text (e.g. para beginning 142) is too brief to be certain how it was applied. However, table s7 reveals that each location was modeled separately – hence cultivar and site will be confounded.

One limitation of the "production value" approach is that it fails to reflect that farmers would be expected to adopt alternative strategies rather than persist in spite of crop failures. Alternatives to insect pollination include, for example, breeding and adoption of self-pollinating (or apomictic) varieties or the use of mechanical pollen spraying. A better estimate of the economic value of crop pollination is then the difference in cost of the two strategies (ie a substitution cost). It would be helpful for the discussion to explain how the economic value estimated here should be interpreted.

The y axis label in fig S4 should be labeled "estimated pollen deposition" because actual deposition was not measured in this study. The tight relationship between x and y is in large part a consequence of the method, because PPV values are applied as constants. Real pollen counts would of course be much more variable.

Review form: Reviewer 2

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?
N/A

Is it adequate?
N/A

Do you have any ethical concerns with this paper?
No

Comments to the Author
General comments
In their study "Crop yield in the USA is frequently limited by a lack of
Pollinators" Reilly et al. assessed the extent to which crop yield is of seven crops across several major production regions of the USA is limited by suboptimal insect pollination. Moreover, the authors estimated the nation-wide annual production value of pollination services to these crops, and the relative contribution of wild bees and managed honeybees. Such assessments are important and needed. Despite previous attempts to address such questions, the present study

clearly improves our understanding of the contribution of wild and managed pollinators to crop yield. The sampling design and analyses are convincing to me appropriately performed. I liked the piecewise regression approach used in some of the models to assess pollination limitation. I appreciate that authors also provide additional analyses (complementary/ alternative analyses, sensitivity analyses) and results in the supplementary information.
What I miss in the current version of the manuscript is a more critical presentation and discussion some of the limitations and uncertainties almost inevitably associated with such assessments. Such uncertainties arise for example from the lack data for some components in the analyses. For example, it is almost impossible, but probably also not needed, to get robust per-visit efficiency data for all wild bee species in a crop pollinator community. Table S2 shows that also in this study for some crops such data was unavailable for some species and species groups. I do not think that this invalidates the analyses and respective results, but I find it important that it becomes clear on what basis in terms of available data (including sample sizes) these values were estimated and that the uncertainties involved in such an approach should be explicitly and critically discussed in the main text of the manuscript, and conclusions made with appropriate care. See also specific comments to this point below. Also, there are, inevitably, uncertainties in the estimations of the contribution of pollination services to the annual production values, and extrapolations from the studied sites to the entire nation. For example, the data provided by Klein et al. 2007 about the degree of pollinator dependence used for these estimations are probably the best we have, but we also know that many factors such as crop cultivar etc. affect such values and that many uncertainties are associated with them. Also, I very much appreciate the sensitivity analyses provided in the supplementary information presenting valuations before and after subtracting variable production costs. Such information helps to better understand the estimates and potential uncertainties. As these analyses show, they can be substantial. In conclusion, I find it important to include such points (see also specific comments below) in a more critical discussion of the limitations/uncertainties involved in such assessments in the main text of the manuscript. I also appreciate the presentation of results if the alternative hand-pollination approach was used to assess pollination limitation of crop yield in blueberry. I would encourage the authors to consider to discuss these results also in the main manuscript, maybe together with a discussion of particular strengths and potential limitations of the two different approaches, which I think could be valuable for readers (see specific comments below). I hope find the specific comments below useful.

Specific comments

Abstract
L29: I would explicitly write “insect pollination” instead of pollination
L32: Consider to use “estimate” or “assess” rather than “measure”
L34: Comparable is rather vague. Could you give more precise (quantitative) information here? Also, contributions differ substantially across crop species according to the results, which should become more clear here, I think.
Agriculturally intensive regions: can you provide in a supplementary information about management intensity of crops of studied sites, information about intensity at landscape-scale to better understand the systems studied here?
L35: please clarify that annual production values were estimated here.
L37: I would delete “continued”. We simply know for sure that some species are declining in some regions, or?
L39: Consider to include “… contribute substantially to pollination of most study crops in…” or something similar (e.g. no contribution to almond pollination in the studied region apparently)

Introduction
L44: Can you provide a reference, e.g. IPBES 2016 report?
L52-53: There are some studies though in addtion to the cited ones, e.g.:
Garibaldi, L. A., Carvalheiro, L. G., Vaissière, B. E., Gemmill-Herren, B., Hipólito, J., Freitas, B. M., ... & An, J. (2016). Mutually beneficial pollinator diversity and crop yield outcomes in small and large farms. Science, 351(6271), 388-391.

Garratt, M. P., Bishop, J., Degani, E., Potts, S. G., Shaw, R. F., Shi, A., & Roy, S. (2018). Insect pollination as an agronomic input: Strategies for oilseed rape production. Journal of applied ecology, 55(6), 2834-2842;
Fijen, T. P., Scheper, J. A., Boom, T. M., Janssen, N., Raemakers, I., & Kleijn, D. (2018). Insect pollination is at least as important for marketable crop yield as plant quality in a seed crop. Ecology letters, 21(11), 1704-1713.
L71: but see e.g. Garratt et al. 2018 or Fijen et al. 2018 mentioned above…
L73-74: consider to mention also other managed bees
L76: For me Kleijn et al. 2015 is missing here (cited somewhere else in the manuscript)

Methods
L116: Could you provide quantitative information supporting the statements made here and in the next sentences (e.g. in a supplementary information)?
L134-138: Is fruit set and fruit weight more or less equal to crop yield for all studied crops? Or can you briefly justify the use of these two variable as a proxy of crop yield?
L142: For me it is confusing that sometimes pollen limitation is used, while in other parts of the ms pollination limitation is used. Maybe I miss something here, but I suggest to consistently use pollination limitation throughout, unless the authors have clear reasons to use pollen limitation in some sections.
L142: Could pollination limitation also vary within sites?
L155: Per-visit efficiency: see general comment above and specific comments below. Pollen deposition rates are probably a good proxy for per-visit efficiency. Consider to discuss, however, that also other factors, such as the proportion of outcross pollen, or in some crops such as apple, the proportion of pollinizer pollen could also be relevant.

Results
L189: Here you give the proportion of pollination limitation in % of sites (which corresponds to the description of this analysis in the methods section (L142-149). In the Supplementary Methods (L13-16) I understand that estimates were calculated for each transect. Did you assume pollination limitation for a site if you found indication for this for all? More half? At least one? Transect of a certain site? Or do I misunderstand something here? Could you please clarify.
L190: Delete bracket
L192-196: Consider to move to discussion section and to provide refs for some of the statements made
L214-226: I think also the most important results of the analyses subtracting variable cost should be mentioned here. Further, the large difference in outcomes depending on whether variable costs are subtracted or not should be discussed in the discussion section of the main manuscript, in my opinion. Currently, this is somewhat hidden in the supplementary.
L236: other papers could be cited (see above)
L247: but very variable across crops and regions. I think this deserves mentioning here.

L248-250: In my opinion this statement is too strong and a critical statement about lack of robust efficiency data for most wild bee species/taxa in the community for example in the studied cherry, apple or also pumpkin systems is needed considering the available data presented in Table S2.

Supplementary methods and analyses
General comment: If possible (with respect to the world limit for PRSB) I would encourage authors to move some of the most important parts of the supplementary methods to the main manuscript. For me, the description of methods and analyses was a difficult to follow, as it is quite scattered across main text and several supplementary sections. I would like to emphasize that I very much appreciate that authors have also tested additional analyses and provide detailed descriptions of all these analyses, which helps to understand methods and results, which certainly makes the paper stronger.

L23: “Tart cherry in Pennsylvania was not included in this analysis due to insufficient data”. Please provide more detailed information (quantitative data) here clarifying why tart cherry could not be included in these analyses.
L187: Supplementary hand pollination (flowers receiving pollen from hand-pollination on top of the pollen they receive by pollinators; flowers not bagged) or hand pollination of bagged flowers?
L191: Why only berry weight? What about fruit set?
L198-207: I think it is a strength of this study that you assess pollination limitation also with an alternative method (using a hand pollination approach). Different methods have their particular strengths and potential limitations. For, example, focusing on yield is probably of more direct relevance to growers. On the other hand, estimates of pollination limitation following this approach may be somewhat less precise and more correlational, as many other factors (mentioned by authors) affect crop yield, not only pollination services. The hand pollination approach on the other hand could provide a more direct and potentially more precise method to estimate pollination limitation, but it might be less directly relevant to farmers and final crop yield. I think mentioning the hand pollination results and discussing the advantages and disadvantages of both approaches could be valuable for readers (see also general comment above).

L58-62: It would be helpful to included sample sizes in Table S2, especially for the per-visit efficiency data. Table S2 shows that efficiency data were not available for all non-Apis bee species groups, and probably only a fraction of the species in these groups. It would be almost impossible, and probably also not needed to have such data available for all wild bee species of a system. Deriving a single value from a very broad wild bee group including a large number of species differing in traits likely affecting pollination efficiency, however, likely ignores considerable within-group variation in efficiency. I do not think this lack of data invalidates such analyses and outcomes, but I find it important that it becomes very clear on what basis in terms of available data these values were estimated and that the limitations und large uncertainties involved in such an approach should be explicitly discussed in the main text of the manuscript. In particular assuming honeybee efficiencies for many wild bee species/species groups due to a lack of data and using efficiency values estimated for apple in cherry could be problematic and should at least be critically discussed. Some efficiency data for Sweet cherry pollination would be available for a European region (Eeraerts, M., Vanderhaegen, R., Smagghe, G., & Meeus, I. (2019). Pollination efficiency and foraging behaviour of honey bees and non-Apis bees to sweet cherry. Agricultural and Forest Entomology).

L83-94: This paragraph is identical to such a paragraph in the main text…
L115-122: I appreciate this discussion of potential uncertainties and limitations of such extrapolations here.
L161-165: In my opinion the strong effect of whether or not to subtract variable costs from crop production values should be mentioned and discussed in the main text (see general comment).

Decision letter (RSPB-2019-2393.R0)

12-Dec-2019

Dear Dr Reilly:

I am writing to inform you that your manuscript RSPB-2019-2393 entitled "Crop yield in the USA is frequently limited by a lack of pollinators" has, in its current form, been rejected for publication in Proceedings B.

This action has been taken on the advice of referees, who have recommended that substantial revisions are necessary. With this in mind 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.

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.

Please note that this decision may (or may not) have taken into account confidential comments.

In your revision process, please take a second look at how open your science is; our policy is that all data involved with the study should be made openly accessible-- see:
https://royalsociety.org/journals/ethics-policies/data-sharing-mining/
Insufficient sharing of data can delay or even cause rejection of a paper.

Sincerely,
Professor John Hutchinson, Editor
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
The manuscript assesses the extent of pollination limitation in seven crops and 131 locations throughout the USA, additionally quantifying the relative contributions of honebees and wild pollinators to crop pollination. The two reviewers of the manuscript appreciate the work the authors have conducted and think that it has the potential to represent a valuable contribution to the literature. At the same time, they both had several caveats about the study, some of them substantial. Reviewer 1 had reviewed a previous version of the manuscript for another journal (PNAS). By comparing the two versions and referring to the original review made for PNAS, this reviewer was surprised to see that some of the suggestions made in the previous review may not have been incorporated or were done in a way that generates some additional concerns. One of the concerns regards the methods for assessing apple fruit set, specifically regarding the timing of the pollination assessments relative to the timing of thinning. According to this reviewer, the manuscript submitted to PNAS stated that these studies had been done before thinning, whereas the current manuscript states that they were done after thinning, but the data reported in the relevant figure (Fig. S2) appear to be the same. I was very concerned about this issue, and I think the authors should offer a convincing explanation of this apparent inconsistency. Reviewer 1 also pointed out that the limited sampling effort in many plots is quite small, which may affect estimates of pollination limitation, leading to (wrongly) concluding that there is no pollination

limitation when in fact it may be present. This reviewer (as well as Reviewer 2) also points out that the authors are confounding pollination limitation and pollen limitation, which in agro-ecosystems are known not be the same, and that referring to "yield" may be incorrect in the context of this study, as only two of the seven crops were assessed on a per-area basis, which is necessary for correctly inferring yield. Reviewer 1 makes a number of additional comments about other aspects of the manuscript. In turn, reviewer 2 thinks that the manuscript should do a better job at discussing the limitations and uncertainties of the study, especially regarding limited data for some of the analyses. This reviewer makes a number of more specific but still important suggestions throughout the manuscript, which should also be considered and dealt with by the authors.

In addition to the comments made by the two reviewers, I have one relatively minor comment about Fig. 1. I think it may not be clear for the readers what precisely the authors mean in panel C when they say that pollination is limiting in some places. The truncated line in this panel may indicate simply that crop yield saturates at some level of pollinator visits, so that pollination is limited up to some level, beyond which it is no longer pollinator limited presumably because of the saturation of pollination. Whether this saturation occurs in some places or in all of them seems irrelevant, or at least does not follow clearly from the figure. I think the authors should clarify this point in the introduction and in the figure legend, and if there is an inherently spatial component to the concepts behind this figure it should be clearly explained. I know this may be clear once the reader learns about the study design in the methods section, but at the point of the introduction when the figure is introduced I think this panel may be confusing.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
I realized when I started to review this submission that it was a version of a paper that I previously reviewed at another journal (PNAS) and where I raised quite a number of concerns. Now that I examine the ms for review I realise that many of my concerns relate not just to this version, but to changes that have been made (and many that have not been made) since the last version that I saw. In particular I have picked up a change that requires some explanation.

The original ms had the following text regarding methods for assessing apple fruit set
"Although apples are typically thinned to achieve fruit that meet fresh-market standards, all apple pollination assessments were **before** this to avoid the variation among farms in their thinning practices." (emphasis added)
This prompted me to raise the following concern in my original review
"Getting a true yield measure is difficult ……..This problem is greatest for apples, where on line 206 it is explained that fruit set was assessed before the crop was thinned. Assessing fruit set this way is certainly more likely to detect pollen limitation than a post thinning assessment, but the response variable is not commercial yield. It is quite possible that, for example there is a significant relationship between flower visits and early fruit set (such as was found, line 788) but no relationship with yield of mature fruit."
This revised ms has text in the methods that indicates a change since the PNAS version, perhaps in response to the concern that I raised. Here is the text:
"Apples are typically thinned to achieve fruit that meet fresh-market standards; thus our apple fruit counts were taken **post-thinning** to be more directly related to harvestable yield. This is a conservative approach, because post-thinning measurements are less likely than those taken pre-thinning to detect the effect of pollination limitation." (emphasis added).
I was surprised to see this, because it suggested that both pre-thinning and post-thinning assessment were done, and therefore the authors were able to re-analyse using a revised data set. Then I examined the figures (fig s2) and noted that the data were exactly the same as that

presented in the earlier version, even though the methods indicated a change in approach. I can think of at three possible explanations for this:
1) It may be that the first version I read was in fact wrong with regard the method and therefore it is possible that this revised version is correct, whereas the earlier version was wrong.
2) It may be that the authors made a mistake in submitting the old figures in the revised ms.
3) It may be that the authors changed the method text but failed to run a new analysis.

Other concerns raised in my original review, and which remain a concern
Table S3 reveals that the sampling effort for flower observations is quite small in many cases. In a number of crops observations were only made on one date per year and at only one time of day. For a number of crops the total time spent observing in each transect was only 6.7 minutes. Given that pollinators are active for many hours of each day, and flowering for most crops extends over more than a week, one would have to expect that this level of sampling will have limited power to detect patterns. Adding more noise, it is known that not all visits are equal (in terms of how the flower is handled by bees, for example).
Because of the difficulty in getting comprehensive observations on visitor frequency over the life of flowers or the full flowering season it is not unexpected that pollen limitation might be detected by other methods (like distance from hive effects and pollen supplementation effects) in that same sites that flower visitor observations fail to detect a pattern. Other studies have shown this. It is certainly apparent that even where positive relationships are detected in this study, they are generally not tight fits (ie large y residuals in the graphs of fig s2).
This problem is likely to lead to conservative conclusions, e.g. they are likely to fail to detect pollen limitation even when it is present. It is also expected that the relatively poor fit of the relationships also makes estimates of the "breakpoint" (i.e. the threshold in relationships that first increase and then flatten) is likely to have wide error margins. The analytical methods do not explore the consequences of this problem.
In the paragraph beginning line 142 (and then throughout) the paper refers to the idea that the study assesses pollen limitation, when in fact the method measures visitor limitation. There is an unstated assumption that the two are almost the same, but in agricultural systems we know that pollen limitation can also be a consequence of planting design especially in crops that have self-infertile varieties which require cross pollination between different rows.
Other studies have explicitly examined the importance of pollinator diversity in reducing likelihood of pollen limitation in two quite different ways, i.e. spatially and temporally. This study does not to examine the temporal component (no exploration of time of day, time of season or year effects). The variability examined is therefore only spatial (transect to transect and site to site). The text should be more explicit about this design decision and the consequences.
The paper uses the term "yield" throughout, which is generally defined as the mass of the crop produced per unit area of farmed land. It is a powerful variable because it relates very closely to efficiency of production. Because one of the aims of the study is to understand the economic value of crop pollination, it is particularly important to use measures of production that relate closely to commercial outcomes. Across the different production metrics in this study only two crops (Table s4, pumpkin and melon) were assessed on a per area basis. For most crops in the study the measures were fruit per flower or branch, which are well short of true yield. Getting a true yield measure is difficult in the experimental context, so this is not surprising. Nevertheless it is important to be careful in the choice of words.
Other studies show that there can be large differences in breeding system (and therefore vulnerability to pollen limitation) among cultivars or varieties of the same crop. However, the method description is not clear regarding how differences in cultivar are handled, except the comment on line 29 that implies the "regionally dominant" cultivar was examined. The description of the model fitting in the text (e.g. para beginning 142) is too brief to be certain how it was applied. However, table s7 reveals that each location was modeled separately – hence cultivar and site will be confounded.

One limitation of the "production value" approach is that it fails to reflect that farmers would be expected to adopt alternative strategies rather than persist in spite of crop failures. Alternatives to insect pollination include, for example, breeding and adoption of self-pollinating (or apomictic)

varieties or the use of mechanical pollen spraying. A better estimate of the economic value of crop pollination is then the difference in cost of the two strategies (ie a substitution cost). It would be helpful for the discussion to explain how the economic value estimated here should be interpreted.

The y axis label in fig S4 should be labeled "estimated pollen deposition" because actual deposition was not measured in this study. The tight relationship between x and y is in large part a consequence of the method, because PPV values are applied as constants. Real pollen counts would of course be much more variable.

Referee: 2

Comments to the Author(s)
General comments
In their study "Crop yield in the USA is frequently limited by a lack of
Pollinators" Reilly et al. assessed the extent to which crop yield is of seven crops across several major production regions of the USA is limited by suboptimal insect pollination. Moreover, the authors estimated the nation-wide annual production value of pollination services to these crops, and the relative contribution of wild bees and managed honeybees. Such assessments are important and needed. Despite previous attempts to address such questions, the present study clearly improves our understanding of the contribution of wild and managed pollinators to crop yield. The sampling design and analyses are convincing to me appropriately performed. I liked the piecewise regression approach used in some of the models to assess pollination limitation. I appreciate that authors also provide additional analyses (complementary/ alternative analyses, sensitivity analyses) and results in the supplementary information.
What I miss in the current version of the manuscript is a more critical presentation and discussion some of the limitations and uncertainties almost inevitably associated with such assessments. Such uncertainties arise for example from the lack data for some components in the analyses. For example, it is almost impossible, but probably also not needed, to get robust per-visit efficiency data for all wild bee species in a crop pollinator community. Table S2 shows that also in this study for some crops such data was unavailable for some species and species groups. I do not think that this invalidates the analyses and respective results, but I find it important that it becomes clear on what basis in terms of available data (including sample sizes) these values were estimated and that the uncertainties involved in such an approach should be explicitly and critically discussed in the main text of the manuscript, and conclusions made with appropriate care. See also specific comments to this point below. Also, there are, inevitably, uncertainties in the estimations of the contribution of pollination services to the annual production values, and extrapolations from the studied sites to the entire nation. For example, the data provided by Klein et al. 2007 about the degree of pollinator dependence used for these estimations are probably the best we have, but we also know that many factors such as crop cultivar etc. affect such values and that many uncertainties are associated with them. Also, I very much appreciate the sensitivity analyses provided in the supplementary information presenting valuations before and after subtracting variable production costs. Such information helps to better understand the estimates and potential uncertainties. As these analyses show, they can be substantial. In conclusion, I find it important to include such points (see also specific comments below) in a more critical discussion of the limitations/uncertainties involved in such assessments in the main text of the manuscript. I also appreciate the presentation of results if the alternative hand-pollination approach was used to assess pollination limitation of crop yield in blueberry. I would encourage the authors to consider to discuss these results also in the main manuscript, maybe together with a discussion of particular strengths and potential limitations of the two different approaches, which I think could be valuable for readers (see specific comments below). I hope find the specific comments below useful.

Specific comments

Abstract
L29: I would explicitly write "insect pollination" instead of pollination
L32: Consider to use "estimate" or "assess" rather than "measure"
L34: Comparable is rather vague. Could you give more precise (quantitative) information here? Also, contributions differ substantially across crop species according to the results, which should become more clear here, I think.
Agriculturally intensive regions: can you provide in a supplementary information about management intensity of crops of studied sites, information about intensity at landscape-scale to better understand the systems studied here?
L35: please clarify that annual production values were estimated here.
L37: I would delete "continued". We simply know for sure that some species are declining in some regions, or?
L39: Consider to include "… contribute substantially to pollination of most study crops in…" or something similar (e.g. no contribution to almond pollination in the studied region apparently)

Introduction
L44: Can you provide a reference, e.g. IPBES 2016 report?
L52-53: There are some studies though in addtion to the cited ones, e.g.:
Garibaldi, L. A., Carvalheiro, L. G., Vaissière, B. E., Gemmill-Herren, B., Hipólito, J., Freitas, B. M., ... & An, J. (2016). Mutually beneficial pollinator diversity and crop yield outcomes in small and large farms. Science, 351(6271), 388-391.
Garratt, M. P., Bishop, J., Degani, E., Potts, S. G., Shaw, R. F., Shi, A., & Roy, S. (2018). Insect pollination as an agronomic input: Strategies for oilseed rape production. Journal of applied ecology, 55(6), 2834-2842;
Fijen, T. P., Scheper, J. A., Boom, T. M., Janssen, N., Raemakers, I., & Kleijn, D. (2018). Insect pollination is at least as important for marketable crop yield as plant quality in a seed crop. Ecology letters, 21(11), 1704-1713.
L71: but see e.g. Garratt et al. 2018 or Fijen et al. 2018 mentioned above…
L73-74: consider to mention also other managed bees
L76: For me Kleijn et al. 2015 is missing here (cited somewhere else in the manuscript)

Methods
L116: Could you provide quantitative information supporting the statements made here and in the next sentences (e.g. in a supplementary information)?
L134-138: Is fruit set and fruit weight more or less equal to crop yield for all studied crops? Or can you briefly justify the use of these two variable as a proxy of crop yield?
L142: For me it is confusing that sometimes pollen limitation is used, while in other parts of the ms pollination limitation is used. Maybe I miss something here, but I suggest to consistently use pollination limitation throughout, unless the authors have clear reasons to use pollen limitation in some sections.
L142: Could pollination limitation also vary within sites?
L155: Per-visit efficiency: see general comment above and specific comments below. Pollen deposition rates are probably a good proxy for per-visit efficiency. Consider to discuss, however, that also other factors, such as the proportion of outcross pollen, or in some crops such as apple, the proportion of pollinizer pollen could also be relevant.

Results
L189: Here you give the proportion of pollination limitation in % of sites (which corresponds to the description of this analysis in the methods section (L142-149). In the Supplementary Methods (L13-16) I understand that estimates were calculated for each transect. Did you assume pollination limitation for a site if you found indication for this for all? More half? At least one? Transect of a certain site? Or do I misunderstand something here? Could you please clarify.
L190: Delete bracket

L192-196: Consider to move to discussion section and to provide refs for some of the statements made
L214-226: I think also the most important results of the analyses subtracting variable cost should be mentioned here. Further, the large difference in outcomes depending on whether variable costs are subtracted or not should be discussed in the discussion section of the main manuscript, in my opinion. Currently, this is somewhat hidden in the supplementary.
L236: other papers could be cited (see above)
L247: but very variable across crops and regions. I think this deserves mentioning here.

L248-250: In my opinion this statement is too strong and a critical statement about lack of robust efficiency data for most wild bee species/taxa in the community for example in the studied cherry, apple or also pumpkin systems is needed considering the available data presented in Table S2.

Supplementary methods and analyses
General comment: If possible (with respect to the world limit for PRSB) I would encourage authors to move some of the most important parts of the supplementary methods to the main manuscript. For me, the description of methods and analyses was a difficult to follow, as it is quite scattered across main text and several supplementary sections. I would like to emphasize that I very much appreciate that authors have also tested additional analyses and provide detailed descriptions of all these analyses, which helps to understand methods and results, which certainly makes the paper stronger.

L23: "Tart cherry in Pennsylvania was not included in this analysis due to insufficient data". Please provide more detailed information (quantitative data) here clarifying why tart cherry could not be included in these analyses.
L187: Supplementary hand pollination (flowers receiving pollen from hand-pollination on top of the pollen they receive by pollinators; flowers not bagged) or hand pollination of bagged flowers?
L191: Why only berry weight? What about fruit set?
L198-207: I think it is a strength of this study that you assess pollination limitation also with an alternative method (using a hand pollination approach). Different methods have their particular strengths and potential limitations. For, example, focusing on yield is probably of more direct relevance to growers. On the other hand, estimates of pollination limitation following this approach may be somewhat less precise and more correlational, as many other factors (mentioned by authors) affect crop yield, not only pollination services. The hand pollination approach on the other hand could provide a more direct and potentially more precise method to estimate pollination limitation, but it might be less directly relevant to farmers and final crop yield. I think mentioning the hand pollination results and discussing the advantages and disadvantages of both approaches could be valuable for readers (see also general comment above).

L58-62: It would be helpful to included sample sizes in Table S2, especially for the per-visit efficiency data. Table S2 shows that efficiency data were not available for all non-Apis bee species groups, and probably only a fraction of the species in these groups. It would be almost impossible, and probably also not needed to have such data available for all wild bee species of a system. Deriving a single value from a very broad wild bee group including a large number of species differing in traits likely affecting pollination efficiency, however, likely ignores considerable within-group variation in efficiency. I do not think this lack of data invalidates such analyses and outcomes, but I find it important that it becomes very clear on what basis in terms of available data these values were estimated and that the limitations und large uncertainties involved in such an approach should be explicitly discussed in the main text of the manuscript. In particular assuming honeybee efficiencies for many wild bee species/species groups due to a lack of data and using efficiency values estimated for apple in cherry could be problematic and should at least be critically discussed. Some efficiency data for Sweet cherry pollination would be available for a European region (Eeraerts, M., Vanderhaegen, R., Smagghe, G., & Meeus, I. (2019).

Pollination efficiency and foraging behaviour of honey bees and non-Apis bees to sweet cherry. Agricultural and Forest Entomology).

L83-94: This paragraph is identical to such a paragraph in the main text…
L115-122: I appreciate this discussion of potential uncertainties and limitations of such extrapolations here.
L161-165: In my opinion the strong effect of whether or not to subtract variable costs from crop production values should be mentioned and discussed in the main text (see general comment).

Author's Response to Decision Letter for (RSPB-2019-2393.R0)

See Appendix A.

RSPB-2020-0922.R0

Review form: Reviewer 1

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?
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.
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
The authors have thoughtfully and appropriately responded to most of the comments made in the first review, and the manuscript is now very much improved.
My primary on-going concern is with regards use of the word 'yield'. In my original review I commented as follows:

Across the different production metrics in this study only two crops (Table s4, pumpkin and melon) were assessed on a per area basis. For most crops in the study the measures were fruit per flower or branch, which are well short of true yield. Getting a true yield measure is difficult in the experimental context, so this is not surprising. Nevertheless it is important to be careful in the choice of words.

To which the authors responded:

13. We agree. Clarification added on lines 161-165.

However, the lines numbered 161-165 do not relate to this problem, so this I an error.
I think the actual words of clarification are at lines 146-148, which has the following text:

Thus our use of the term yield necessarily differs somewhat by crop, but matches commonly used proxies for yield (Garibaldi et al. 2013, Bartomeus et al. 2014).

In my view this text is not sufficient to solve the problem. It remains the case that the word yield appears 55 times in the main document and 18 times in the ESM. Most significantly it appears in the title and the abstract. The reader has to dig quite deep to establish that for most crops in the study what is actually measured is a proxy for yield, and that no assessment is made regarding how effective the proxy measure is. This matters because the focus of the study is on economic benefits for agriculture, and in agricultural terms yield is understood as a per land area measure. In some circumstances one might examine the yield relative to another high cost input, such as irrigation water or fertilizer. But looking at 'yield' on a per flower or per branch basis is a biology perspective that is poorly linked to any economic measure.

I do not consider this a hard problem to solve. When referring to measures such as fruit per flower or fruit per branch, replace the word yield with the less specific "production". Some uses of the word yield are appropriate because they are made in reference to theory, or to other studies, or to the crops where production was measured on a per area basis (i.e. watermelon, pumpkin). But when referring to the measures such as fruit per flower or fruit per branch, replace the word yield with the less specific "production".

Additional minor comments
I suggest inserting the following recent paper (from Proc Roy Soc) which is relevant at line 80
"… , Rader et al. 2016) and that the diversity of wild bee visitors is higher when crops are grown in their biogeographic region of origin (Brown and Cunningham 2019)."
Brown J, Cunningham SA 2019 Global scale drivers of crop visitor diversity and the historical development of agriculture Proceedings of the Royal Society B 286, 20192096

Regarding the following response to reviewers :
14. Yes, cultivar and site are confounded. From our perspective, this is not necessarily undesirable…

The explanation given is a fair one, but this thinking is not reflected in the paper. It would be appropriate to mention this confounding in the supplementary methods.

Decision letter (RSPB-2020-0922.R0)

24-May-2020

Dear Dr Reilly:

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 John Hutchinson, Editor
mailto: proceedingsb@royalsociety.org

Associate Editor Board Member
Comments to Author:
The authors have done an excellent job at incorporating the suggestions made by the reviewers and by myself, which has clarified and improved several important aspects of the manuscript. One of the reviewers who reassessed the manuscript found that the definition of "yield" is still unclear. This reviewer made a couple of additional minor comments. I thus encourage the authors to consider these additional comments and to submit a revised version of the manuscript incorporating these suggestions.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s).
The authors have thoughtfully and appropriately responded to most of the comments made in the first review, and the manuscript is now very much improved.
My primary on-going concern is with regards use of the word 'yield'. In my original review I commented as follows:

Across the different production metrics in this study only two crops (Table s4, pumpkin and melon) were assessed on a per area basis. For most crops in the study the measures were fruit per flower or branch, which are well short of true yield. Getting a true yield measure is difficult in the experimental context, so this is not surprising. Nevertheless it is important to be careful in the choice of words.

To which the authors responded:

13. We agree. Clarification added on lines 161-165.

However, the lines numbered 161-165 do not relate to this problem, so this I an error.
I think the actual words of clarification are at lines 146-148, which has the following text:

Thus our use of the term yield necessarily differs somewhat by crop, but matches commonly used proxies for yield (Garibaldi et al. 2013, Bartomeus et al. 2014).

In my view this text is not sufficient to solve the problem. It remains the case that the word yield appears 55 times in the main document and 18 times in the ESM. Most significantly it appears in the title and the abstract. The reader has to dig quite deep to establish that for most crops in the study what is actually measured is a proxy for yield, and that no assessment is made regarding how effective the proxy measure is. This matters because the focus of the study is on economic benefits for agriculture, and in agricultural terms yield is understood as a per land area measure. In some circumstances one might examine the yield relative to another high cost input, such as irrigation water or fertilizer. But looking at 'yield' on a per flower or per branch basis is a biology perspective that is poorly linked to any economic measure.

I do not consider this a hard problem to solve. When referring to measures such as fruit per flower or fruit per branch, replace the word yield with the less specific "production". Some uses of the word yield are appropriate because they are made in reference to theory, or to other studies, or to the crops where production was measured on a per area basis (i.e. watermelon, pumpkin). But when referring to the measures such as fruit per flower or fruit per branch, replace the word yield with the less specific "production".

Additional minor comments
I suggest inserting the following recent paper (from Proc Roy Soc) which is relevant at line 80
"… , Rader et al. 2016) and that the diversity of wild bee visitors is higher when crops are grown in their biogeographic region of origin (Brown and Cunningham 2019)."
Brown J, Cunningham SA 2019 Global scale drivers of crop visitor diversity and the historical development of agriculture Proceedings of the Royal Society B 286, 20192096

Regarding the following response to reviewers :
14. Yes, cultivar and site are confounded. From our perspective, this is not necessarily undesirable…
The explanation given is a fair one, but this thinking is not reflected in the paper. It would be appropriate to mention this confounding in the supplementary methods.

Author's Response to Decision Letter for (RSPB-2020-0922.R0)

See Appendix B.

Decision letter (RSPB-2020-0922.R1)

07-Jul-2020

Dear Dr Reilly

I am pleased to inform you that your manuscript entitled "Crop production in the USA is frequently limited by a lack of pollinators" has been accepted for publication in Proceedings B. Congratulations!!

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

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 10 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,
Dr John Hutchinson
Editor, Proceedings B

mailto: proceedingsb@royalsociety.org

Associate Editor:
Board Member
Comments to Author:
I think the authors have successfully incorporated all of the reviewer suggestions. Congratulations to the authors for their excellent work.

Appendix A

12-Dec-2019

Dear Dr Reilly:

I am writing to inform you that your manuscript RSPB-2019-2393 entitled "Crop yield in the USA is frequently limited by a lack of pollinators" has, in its current form, been rejected for publication in Proceedings B.

This action has been taken on the advice of referees, who have recommended that substantial revisions are necessary. With this in mind 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.

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.

Please note that this decision may (or may not) have taken into account confidential comments.

In your revision process, please take a second look at how open your science is; our policy is that all data involved with the study should be made openly accessible-- see: https://royalsociety.org/journals/ethics-policies/data-sharing-mining/ Insufficient sharing of data can delay or even cause rejection of a paper.

Sincerely,

Professor John Hutchinson, Editor
mailto: proceedingsb@royalsociety.org

Associate Editor
Board Member: 1
Comments to Author:
The manuscript assesses the extent of pollination limitation in seven crops and 131 locations throughout the USA, additionally quantifying the relative contributions of honebees and wild pollinators to crop pollination. The two reviewers of the manuscript appreciate the work the authors have conducted and think that it has the potential to represent a valuable contribution to the literature. At the same time, they both had several caveats about the study, some of them substantial.

Reviewer 1 had reviewed a previous version of the manuscript for another journal (PNAS). By comparing the two versions and referring to the original review made for PNAS, this reviewer was surprised to see that some of the suggestions made in the previous review may not have been incorporated or were done in a way that generates some additional concerns.

AUTHORS' RESPONSE

1. Unfortunately since this manuscript was rejected by PNAS, we did not have a chance to respond to this reviewer's comments, but we are glad to have the opportunity now. The main changes we made after receiving the PNAS reviews and before submitting to PRSL are as follows: 1) fixing the incorrect text about apple thinning (see below), 2) numerous clarity-related changes to the introduction, methods, and discussion, 3) expanded our analytical methods for the pollination limitation analysis (described in ESM lines 5-40), 4) added new data, analysis, and text reporting on a second method for testing for pollination limitation, using hand-pollination experiments; we only did these experiments in one crop (blueberry), which is why we had not included them before, but they do add support to the main result that there is pollination limitation in blueberry. See ESM lines 231-257, and new figure S7 and table S9, 5) adding sensitivity analyses for blueberry regions treated separately or combined across regions; see ESM lines 218-228 and new figure S6, 6) adding more explanation about our use of the production value method (see ESM lines 72-160), and 7) incorporating references suggested by the PNAS reviewers.

One of the concerns regards the methods for assessing apple fruit set, specifically regarding the timing of the pollination assessments relative to the timing of thinning. According to this reviewer, the manuscript submitted to PNAS stated that these studies had been done before thinning, whereas the current manuscript states that they were done after thinning, but the data reported in the relevant figure (Fig. S2) appear to be the same. I was very concerned about this issue, and I think the authors should offer a convincing explanation of this apparent inconsistency.

2. We understand the editor's and reviewer's concern here; in your shoes we would have been very concerned as well. The text that the reviewer remembers from the PNAS submission was incorrect, which is why we changed it before submitting to PRSL. The statement about fruit set being collected before thinning was added by a co-author during the manuscript revision, and we (the lead and senior authors) missed it. After we received the PNAS reviews, we noticed the error and corrected the text. Figure S2 was always correct and therefore didn't change between the PNAS submission and the current version (except that there is now an additional row of panels for Florida blueberry that were not present in the PNAS version). In sum, both the text and figure S2 are correct in the current version, and figure S2 was correct in the version submitted to PNAS, but the text was incorrect. The details on how our apple data were collected are as follows: Apple pollinator observations were performed between May 3 and May 22. Chemical thinning is typically applied in June when fruits are very small, at most 30 days after peak bloom. Our apple fruit counts were performed in mid-July, so definitely post-thinning.

Reviewer 1 also pointed out that the limited sampling effort in many plots is quite small, which may affect estimates of pollination limitation, leading to (wrongly) concluding that there is no pollination limitation when in fact it may be present.

3. Although it is certainly possible that sampling effort could influence our results, we do not find this to be of primary concern given that we **did** detect pollination limitation in most systems. The point is that low sample size primarily affects the likelihood of false negatives (i.e. more likely to see no evidence for limitation when it really does exist). The fact that we did find evidence for limitation in a number of crops is consistent with our sample sizes being adequate. In other words, as the reviewer states, the potential noise and undersampling should be a bias against against showing pollination limitation, and hence shouldn't undermine the main findings reported in this manuscript.

Furthermore, we note that some of the crops with lower sampling effort (e.g. apple) had some of the strongest results, while intensively sampled crops like watermelon and pumpkin had very weak or no relationship. As a check of this relationship, we performed a logistic regression on the effect of total observation time on whether crops showed clear evidence of pollination limitation in the analysis. There was no relationship ($p=0.53$).

With respect to the limited time allocated to each transect sample, unfortunately this was a necessary part of our study design. In this study we were aiming to get a national perspective, so we had to balance crop and geographical coverage with in-system replication. We collected data at 113 sites across the whole study, with each site sampled multiple times during crop bloom (which itself is a short window, as little as two weeks). Given the limitations on budget and field technicians, this meant the time spent per sample was shorter.

This reviewer (as well as Reviewer 2) also points out that the authors are confounding pollination limitation and pollen limitation, which in agro-ecosystems are known not be the same, and that referring to "yield" may be incorrect in the context of this study, as only two of the seven crops were assessed on a per-area basis, which is necessary for correctly inferring yield. Reviewer 1 makes a number of additional comments about other aspects of the manuscript. In turn, reviewer 2 thinks that the manuscript should do a better job at discussing the limitations and uncertainties of the study, especially regarding limited data for some of the analyses. This reviewer makes a number of more specific but still important suggestions throughout the manuscript, which should also be considered and dealt with by the authors.

4. Pollination limitation verses pollen limitation versus visitor/pollinator limitation: We agree that precise terminology is helpful here. See response #11 below.
Yield: We agree with the reviewer here. See response #13 below

In addition to the comments made by the two reviewers, I have one relatively minor comment about Fig. 1. I think it may not be clear for the readers what precisely the authors mean in panel C when they say that pollination is limiting in some places. The truncated line in this panel may indicate simply that crop yield saturates at some level of pollinator visits, so that pollination is limited up to some level, beyond which it is no longer pollinator limited presumably because of the saturation of pollination. Whether this saturation occurs in some places or in all of them seems irrelevant, or at least does not follow clearly from the figure. I think the authors should clarify this point in the introduction and in the figure legend, and if there is an inherently spatial component to the concepts behind this figure it should be clearly explained. I know this may be clear once the reader learns about the study design in the methods section, but at the point of the introduction when the figure is introduced I think this panel may be confusing.

5. Your point is a good one. If we understand correctly, the issue is that this figure can be interpreted in two ways: 1) that the x axis numbers are referring to potential counts at a single site, and thus the relationship shows that if pollinators increase, limitation will eventually go away. Or 2) that the x axis numbers are referring to measurements across different sites (each point on such a graph being a site), in which case the relationship shows that some sites have lower yield than others, and that this is at least partly due differences in pollination. Among sites with very high pollination, the relationship disappears suggesting that pollination is not limiting there. Although these two interpretations are very similar, our point deals with interpretation 2, so we have added clarification on lines 67-69.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s)
I realized when I started to review this submission that it was a version of a paper that I previously reviewed at another journal (PNAS) and where I raised quite a number of concerns. Now that I examine the ms for review I realise that many of my concerns relate not just to this version, but to changes that have been made (and many that have not been made) since the last version that I saw. In particular I have picked up a change that requires some explanation.

The original ms had the following text regarding methods for assessing apple fruit set
“Although apples are typically thinned to achieve fruit that meet fresh-market standards, all apple pollination assessments were before this to avoid the variation among farms in their thinning practices.” (emphasis added)
This prompted me to raise the following concern in my original review
“Getting a true yield measure is difficult ……..This problem is greatest for apples, where on line 206 it is explained that fruit set was assessed before the crop was thinned. Assessing fruit set this way is certainly more likely to detect pollen limitation than a post thinning assessment, but the response variable is not commercial yield. It is quite possible that, for example there is a significant relationship between flower visits and early fruit set (such as was found, line 788) but no relationship with yield of mature fruit.”

This revised ms has text in the methods that indicates a change since the PNAS version, perhaps in response to the concern that I raised. Here is the text:
“Apples are typically thinned to achieve fruit that meet fresh-market standards; thus our apple fruit counts were taken post-thinning to be more directly related to harvestable yield. This is a conservative approach, because post-thinning measurements are less likely than those taken pre-thinning to detect the effect of pollination limitation.” (emphasis added).
I was surprised to see this, because it suggested that both pre-thinning and post-thinning assessment were done, and therefore the authors were able to re-analyse using a revised data set. Then I examined the figures (fig s2) and noted that the data were exactly the same as that presented in the earlier version, even though the methods indicated a change in approach. I can think of at three possible explanations for this:
1) It may be that the first version I read was in fact wrong with regard the method and therefore it is possible that this revised version is correct, whereas the earlier version was wrong.
2) It may be that the authors made a mistake in submitting the old figures in the revised ms.
3) It may be that the authors changed the method text but failed to run a new analysis.

6. We apologize for the confusion caused by the mistake in the PNAS version. The reviewer's explanation #1 is correct, i.e. that the PNAS version was wrong and the current version is correct. Please see our explanation to the editor above in response #2.

<u>Other concerns raised in my original review, and which remain a concern</u>
Table S3 reveals that the sampling effort for flower observations is quite small in many cases. In a number of crops observations were only made on one date per year and at only one time of day. For a number of crops the total time spent observing in each transect was only 6.7 minutes. Given that pollinators are active for many hours of each day, and flowering for most crops extends over more than a week, one would have to expect that this level of sampling will have limited power to detect patterns.

7. Please see response #3 above to the editor.

Adding more noise, it is known that not all visits are equal (in terms of how the flower is handled by bees, for example).

8. We do not assume that all bee visits are equal across different species and species groups of bees; rather, we use different values of pollen grains deposited per visit for different species and species groups in each crop. These values are given in table S2 (the PPV column) based on previous studies that have determined these values, along with the identity of the species groups and the sample size (number of single-visit pollination experiments done) for each species group in each crop. Our calculations **do** assume though that all visits by individuals of the same species (or species group) of bee are equal, because we use a mean value within species groups rather than modeling the individual variation. We had to do the analyses this way because data on the distribution of single-visit pollen deposition within bee species was not available for most of our crops.

Because of the difficulty in getting comprehensive observations on visitor frequency over the life of flowers or the full flowering season it is not unexpected that pollen limitation might be detected by other methods (like distance from hive effects and pollen supplementation effects) in that same sites that flower visitor observations fail to detect a pattern. Other studies have shown this.

9. We agree that having multiple independent measures of pollen/pollinator limitation would be ideal. The current version of the manuscript has two forms of independent data for one of our crops, blueberry: 1) visitor frequency observations and 2) pollen supplementation experiments. We found that our pollen supplementation results were consistent with our visitor frequency results, but not actually stronger. In fact, we estimated a smaller fraction of transects to be pollination limited using the pollen supplementation methods (reported on ESM lines 248-257).

It is certainly apparent that even where positive relationships are detected in this study, they are generally not tight fits (ie large y residuals in the graphs of fig s2). This problem is likely to lead to conservative conclusions, e.g. they are likely to fail to detect pollen limitation even when it is present. It is also expected that the relatively poor fit of the relationships also makes estimates of the "breakpoint" (i.e. the threshold in relationships that first increase and then flatten) is likely to have wide error margins. The analytical methods do not explore the consequences of this problem.

10. With respect to type II error, see response #3 above. We agree that the positive relationships are not tight. This is not unexpected given the multitude of factors that might potentially affect yield in a particular location, in addition to bee visitation - for example, soil fertility, watering regime, and pest control. We think it is more likely that the true relationship between bee visitation and yield is simply not very predictive given the difficulty of accounting for all the other factors that are part of agricultural production and affect yield, rather than the noisiness of our field data being primarily a function of poor sample size or suboptimal sampling methods that failed to detect what is in reality a very tight relationship.

In the paragraph beginning line 142 (and then throughout) the paper refers to the idea that the study assesses pollen limitation, when in fact the method measures visitor limitation. There is an unstated assumption that the two are almost the same, but in agricultural systems we know that pollen limitation can also be a consequence of planting design especially in crops that have self-infertile varieties which require cross pollination between different rows.

11. There are 3 terms being discussed here, so we will take a moment to define them as best we can. Unfortunately these terms do not have well-accepted definitions and are often used interchangeably in the literature (see e.g. Wilcock and Neiland 2002):
1) pollen limitation occurs when seed production is less than what could be achieved with added pollen. It is usually measured as the difference between seed production (often fruit set) under open pollination and with pollen supplementation by hand (Knight et al. 2005). This is the most general and widely used term and can include effects of both pollen quantity and quality.
2) pollinator limitation (or visitor limitation) occurs when fruit production is less than what could be achieved if pollinators were more abundant/made more visits. This is closely related to pollen limitation because the mechanism by which pollinators affect fruit production is pollen deposition. In practice it is usually measured in the same way as pollen limitation, but with the assumption that the pollen limitation is due to lack of pollinators. This term matches closely with our methodology.
3) pollination limitation is used very inconsistently, but it is sometimes (e.g. Wilcock and Neiland) used to describe cases when pollination fails due to complete lack of pollen dispersal to the stigma, as opposed to cases when some pollen arrives but it is insufficient (pollen limitation).

A further complication is that any type of limitation could refer to different aspects of fruit production, depending on what is of interest in the study, such as fruit set, fruit weight, fruit quality, etc. In practice, all three of these terms could be used for a study looking at the impact of differences in pollinator abundance (as our study does), since we assume that differences in fruit production are due to differences in pollen deposition by the pollinators. However, we agree that "visitor" or "pollinator" limitation is probably the most precise term in this case, and have changed our terminology to "pollinator limitation" throughout. In the section where we discuss the hand supplementation of pollen in blueberry, we instead use the term "pollen limitation" to match the term used most often in the literature for this kind of experiment.

Other studies have explicitly examined the importance of pollinator diversity in reducing likelihood of pollen limitation in two quite different ways, i.e. spatially and temporally. This study does not to examine the temporal component (no exploration of time of day, time of season or year effects). The variability examined is therefore only spatial (transect to transect and site to site). The text should be more explicit about this design decision and the consequences.

12. We have added clarification on this point in the methods section on line 154 and line 25 in the ESM. We didn't focus on time-of-season effects in our study because we instead controlled those effects; we collected all of our data on monoculture plantings during peak bloom. We did not have sufficient data to perform a meaningful analysis of the effect of year.

The paper uses the term "yield" throughout, which is generally defined as the mass of the crop produced per unit area of farmed land. It is a powerful variable because it relates very closely to efficiency of production. Because one of the aims of the study is to understand the economic value of crop pollination, it is particularly important to use measures of production that relate closely to commercial outcomes. Across the different production metrics in this study only two crops (Table s4, pumpkin and melon) were assessed on a per area basis. For most crops in the study the measures were fruit per flower or branch, which are well short of true yield. Getting a true yield measure is difficult in the experimental context, so this is not surprising. Nevertheless it is important to be careful in the choice of words.

13. We agree. Clarification added on lines 161-165.

Other studies show that there can be large differences in breeding system (and therefore vulnerability to pollen limitation) among cultivars or varieties of the same crop. However, the method description is not clear regarding how differences in cultivar are handled, except the comment on line 29 that implies the "regionally dominant" cultivar was examined. The description of the model fitting in the text (e.g. para beginning 142) is too brief to be certain how it was applied. However, table s7 reveals that each location was modeled separately – hence cultivar and site will be confounded.

14. Yes, cultivar and site are confounded. From our perspective, this is not necessarily undesirable since we are simply intending our data to be a representative sample of what exists in these growing regions, and don't intend to make inferences about particular cultivars or sites. Our goal was to make our sample as representative as possible by spreading our sampling over as many sites and transects as we could. The exception is highbush blueberry in BC, MI, and OR (not in FL) where we know that all samples were of the Bluecrop cultivar (see ESM line 31). On a practical level, early on in study design we originally wanted to standardize cultivar nationwide across our study regions, but given the very large number of cultivars grown for these crops and the necessity of matching a cultivar to a farm's environmental conditions, it was not possible for us to do this.

One limitation of the "production value" approach is that it fails to reflect that farmers would be expected to adopt alternative strategies rather than persist in spite of crop failures. Alternatives to insect pollination include, for example, breeding and adoption of self-pollinating (or apomictic) varieties or the use of mechanical pollen spraying. A better estimate of the economic value of crop pollination is then the difference in cost of the two strategies (ie a substitution cost). It would be helpful for the discussion to explain how the economic value estimated here should be interpreted.

15 Substitution cost analyses generally use honey bees as the substitute for wild bees; however since we are looking at the value of both wild bees and honey bees, we cannot employ such a method. The unknowns associated with the future breeding of new selfing varieties or mechanical pollination technologies are too great for us to consider here. Instead, we are content with interpreting the production value method as a short-term snapshot of economic value (e.g. Calderone 2012). Another strategy that farmers can adopt on a short-term time scale is to abandon the now-unnecessary expenditures that they would normally make in their crops, which are no longer needed after pollination failure (e.g. harvest costs). In other words, the production value method could over-estimate the value of pollination because it doesn't subtract these variable costs. To explore this scenario we repeated the full analysis after subtracting variables costs that farmers could at least potentially remediate in the event of pollination failure. We present this analysis in Supplementary analysis 1. We now also discuss this in the main text on line 117-119.

The y axis label in fig S4 should be labeled "estimated pollen deposition" because actual deposition was not measured in this study. The tight relationship between x and y is in large part a consequence of the method, because PPV values are applied as constants. Real pollen counts would of course be much more variable.

16. Done. Also see interesting note added at the bottom of the figure caption after incorporating the PPV values from the new Eeraerts et al 2019 study.

Referee: 2

Comments to the Author(s)
General comments
In their study "Crop yield in the USA is frequently limited by a lack of
Pollinators" Reilly et al. assessed the extent to which crop yield is of seven crops across several major production regions of the USA is limited by suboptimal insect pollination. Moreover, the authors estimated the nation-wide annual production value of pollination services to these crops, and the relative contribution of wild bees and managed honeybees. Such assessments are important and needed. Despite previous attempts to address such questions, the present study clearly improves our understanding of the contribution of wild and managed pollinators to crop yield. The sampling design and analyses are convincing to me appropriately performed. I liked the piecewise regression approach used in some of the models to assess pollination limitation. I appreciate that authors also provide additional analyses (complementary/ alternative analyses, sensitivity analyses) and results in the supplementary information.

What I miss in the current version of the manuscript is a more critical presentation and discussion some of the limitations and uncertainties almost inevitably associated with such assessments. Such uncertainties arise for example from the lack

data for some components in the analyses. For example, it is almost impossible, but probably also not needed, to get robust per-visit efficiency data for all wild bee species in a crop pollinator community.

17. Yes, it is practically impossible to do these per-visit experiments for all wild species. We are very aware of this because our lab collected the pollen deposition data used in the paper for two of the study crops, watermelon and blueberry (Benjamin 2014, Winfree et al. 2007, 2015). It is not possible to identify bees to species on the wing (see lines 150-151); therefore we need to do the single-visit pollen deposition experiments using bee species groups. Bee species are grouped largely by body size and hairiness, which are the two main predictors of pollen deposition per visit (e.g. Stavert et al. 2016, Willmer and Finlayson 2014). This clarification has been added on lines 151-153 in the main text.

Table S2 shows that also in this study for some crops such data was unavailable for some species and species groups. I do not think that this invalidates the analyses and respective results, but I find it important that it becomes clear on what basis in terms of available data (including sample sizes) these values were estimated and that the uncertainties involved in such an approach should be explicitly and critically discussed in the main text of the manuscript, and conclusions made with appropriate care. See also specific comments to this point below.

18. Sample sizes for all single-visit pollen deposition (PPV) estimates have been added to table S2; see also lines 194-195 in main text.

Also, there are, inevitably, uncertainties in the estimations of the contribution of pollination services to the annual production values, and extrapolations from the studied sites to the entire nation. For example, the data provided by Klein et al. 2007 about the degree of pollinator dependence used for these estimations are probably the best we have, but we also know that many factors such as crop cultivar etc. affect such values and that many uncertainties are associated with them.

19. We have added a qualification about the uncertainties in these values on ESM lines 101-104.

Also, I very much appreciate the sensitivity analyses provided in the supplementary information presenting valuations before and after subtracting variable production costs. Such information helps to better understand the estimates and potential uncertainties. As these analyses show, they can be substantial. In conclusion, I find it important to include such points (see also specific comments below) in a more critical discussion of the limitations/uncertainties involved in such assessments in the main text of the manuscript.

20 We are glad that the reviewer appreciated our sensitivity analyses, which are our preferred way of being transparent about the uncertainty in estimated values. White the sensitivity analyses are still in the ESM in the revised version, we added more reference to/discussion of them in the main text on lines 296-300.

I also appreciate the presentation of results if the alternative hand-pollination approach was used to assess pollination limitation of crop yield in blueberry. I would encourage the authors to consider to discuss these results also in the main manuscript, maybe together with a discussion of particular strengths and potential limitations of the two different approaches, which I think could be valuable for readers (see specific comments below). I hope find the specific comments below useful.

21. We agree that the pollen/pollination limitation experiments are important and that they substantially support the results presented in the main text. These results are now discussed explicitly on lines 250-255 in the main text.

Specific comments

Abstract
L29: I would explicitly write “insect pollination” instead of pollination

22. Done.

L32: Consider to use “estimate” or “assess” rather than “measure”

23. Done.

L34: Comparable is rather vague. Could you give more precise (quantitative) information here? Also, contributions differ substantially across crop species according to the results, which should become more clear here, I think.

Agriculturally intensive regions: can you provide in a supplementary information about management intensity of crops of studied sites, information about intensity at landscape-scale to better understand the systems studied here?

24. We feel that the most relevant information on this point is how our study sites are situated with respect to the main production areas for each crop. This is shown in figure S1. However, to add additional information, we also ran a new GIS analysis that calculates the average percentages of agricultural and natural land cover within various radii of our study farms for each crop system (using the NLCD 2016 database). The results of this analysis are presented in ESM table S10 and are copied below, showing an average of 58% and as much as 92% of the surrounding landscape was agriculture within 1km of the sampled fields, with the vast majority of the remaining land cover being natural habitat.

	percent agriculture				percent natural			
crop	1 km	3 km	5 km	10 km	1 km	3 km	5 km	10 km
almond_ca	80	80	73	67	16	17	22	26
apple_mi	67	61	55	49	29	34	38	43
apple_pa	45	35	35	32	47	58	57	60
blueberry_fl	35	25	21	21	60	69	73	71
blueberry_mi	41	30	29	32	48	55	54	50
blueberry_or	84	77	75	71	9	14	13	13
cherry_sweet_wa	13	18	14	10	81	74	78	76
cherry_tart_mi	66	56	50	45	28	36	42	47
cherry_tart_pa	64	55	50	40	24	34	40	51
pumpkin_pa	52	44	44	41	39	47	47	49
watermelon_ca	92	84	81	72	3	9	13	20
watermelon_fl	58	50	46	39	36	41	46	52

L35: please clarify that annual production values were estimated here.

25. Done.

L37: I would delete “continued”. We simply know for sure that some species are declining in some regions, or?

26. Done.

L39: Consider to include “… contribute substantially to pollination of most study crops in…” or something similar (e.g. no contribution to almond pollination in the studied region apparently)

27. Done.

Introduction
L44: Can you provide a reference, e.g. IPBES 2016 report?

28. Reference added.

L52-53: There are some studies though in addtion to the cited ones, e.g.:
Garibaldi, L. A., Carvalheiro, L. G., Vaissière, B. E., Gemmill-Herren, B., Hipólito, J., Freitas, B. M., ... & An, J. (2016). Mutually beneficial pollinator diversity and crop yield outcomes in small and large farms. Science, 351(6271), 388-391.
Garratt, M. P., Bishop, J., Degani, E., Potts, S. G., Shaw, R. F., Shi, A., & Roy, S. (2018). Insect pollination as an agronomic input: Strategies for oilseed rape production. Journal of applied ecology, 55(6), 2834-2842;
Fijen, T. P., Scheper, J. A., Boom, T. M., Janssen, N., Raemakers, I., & Kleijn, D. (2018). Insect pollination is at least as important for marketable crop yield as plant quality in a seed crop. Ecology letters, 21(11), 1704-1713.

29. References added.

L71: but see e.g. Garratt et al. 2018 or Fijen et al. 2018 mentioned above…

30. Reference added.

L73-74: consider to mention also other managed bees

31. Done.

L76: For me Kleijn et al. 2015 is missing here (cited somewhere else in the manuscript)

32. Reference added.

Methods
L116: Could you provide quantitative information supporting the statements made here and in the next sentences (e.g. in a supplementary information)?

33. See response #24 above.

L134-138: Is fruit set and fruit weight more or less equal to crop yield for all studied crops? Or can you briefly justify the use of these two variable as a proxy of crop yield?

34. We used fruit weight as a proxy for yield when it was available (pumpkin, California watermelon, sweet cherry, blueberry) and fruit set or number of fruit otherwise (almond, tart cherry, apple, Florida watermelon). Fruit set or number and fruit weight are commonly used proxies for yield in the pollination literature (e.g. Garibaldi et al. 2013, Bartomeus et al. 2014). See lines 160-168.

L142: For me it is confusing that sometimes pollen limitation is used, while in other parts of the ms pollination limitation is used. Maybe I miss something here, but I suggest to consistently use pollination limitation throughout, unless the authors have clear reasons to use pollen limitation in some sections.

35. We now use "pollinator limitation" throughout the manuscript, except for the supplementation experiment. See response #11 above.

L142: Could pollination limitation also vary within sites?

36. Yes, it potentially could, but that is outside the scope of our study. Our primary goal here is larger-scale effects such as those between regions and crops.

L155: Per-visit efficiency: see general comment above and specific comments below. Pollen deposition rates are probably a good proxy for per-visit efficiency. Consider to discuss, however, that also other factors, such as the proportion of outcross pollen, or in some crops such as apple, the proportion of pollinizer pollen could also be relevant.

37. The reviewer makes a good point here, but given that we are already at the length limit we did not add text on the potential variability due to these factors.

Results
L189: Here you give the proportion of pollination limitation in % of sites (which corresponds to the description of this analysis in the methods section (L142-149). In the Supplementary Methods (L13-16) I understand that estimates were calculated for each transect. Did you assume pollination limitation for a site if you found indication for this for all? More half? At least one? Transect of a certain site? Or do I misunderstand something here? Could you please clarify.

38. Everything was done at the transect level, so "site" would not be correct on line 189 (now line 228), and it should simply say % of transects (which it now does). Specifically, in the analysis, we assume pollination limitation for all transects when a positively-sloped linear model was selected, and we assume no pollination limitation when a no-relationship model or a linear model with a negative slope was selected. When a segmented relationship was selected, we assumed all transects before the estimated breakpoint (i.e. in the part of the graph with a positive slope) were pollination limited, and all transects after the breakpoint (i.e. in the flat part of the graph) were not.

L190: Delete bracket

39. Done.

L192-196: Consider to move to discussion section and to provide refs for some of the statements made

40 We didn't expand on this point or add a reference because we are already at the limit for the number of references, and we think it will be intuitive to most readers that chemical thinning would introduce noise as to the effect of pollinators. A good reference for the pre-thinning measurements being used to measure pollination in apple would be Park et al. 2016.

L214-226: I think also the most important results of the analyses subtracting variable cost should be mentioned here. Further, the large difference in outcomes depending on whether variable costs are subtracted or not should be discussed in the discussion section of the main manuscript, in my opinion. Currently, this is somewhat hidden in the supplementary.

41. We agree with the reviewer here. We have added a discussion of the variable cost analysis on lines 296-300, that points interested readers to the supplemental analysis. Unfortunately we could not add more discussion than this in the main text because we are already at the length limit.

L236: other papers could be cited (see above)

42. Done.

L247: but very variable across crops and regions. I think this deserves mentioning here.

43. Added on line 322.

L248-250: In my opinion this statement is too strong and a critical statement about lack of robust efficiency data for most wild bee species/taxa in the community for example in the studied cherry, apple or also pumpkin systems is needed considering the available data presented in Table S2.

44. The reviewer refers to our statement that "in almost every crop we investigated, the community of wild bees was composed of species with higher than average pollination efficiency (i.e. higher pollen-per-visit values) than that of honey bees." We agree that the sentence was a bit hard to follow, but the basic point was correct. The table below shows the factor by which wild bee species exceeded the honey bee in terms of pollen grains deposited per flower visit. If the average across wild bee groups was equal to the honey bee, we would get a value of 1.0. We have changed this sentence to 'In all six crops we studied, the wild bee species, on average, deposited more pollen per visit than did the honey bee, by a factor of 1.4 to 3.2' (lines 322-324).

Eastern Watermelon	3.2
Western Watermelon	2.2
Pumpkin	1.7
Sweet Cherry	1.5
Blueberry	1.4
Apple	2.0

Supplementary methods and analyses
General comment: If possible (with respect to the world limit for PRSB) I would encourage authors to move some of the most important parts of the supplementary methods to the main manuscript. For me, the description of methods and analyses was a difficult to follow, as it is quite scattered across main text and several supplementary sections.
I would like to emphasize that I very much appreciate that authors have also tested additional analyses and provide detailed descriptions of all these analyses, which helps to understand methods and results, which certainly makes the paper stronger.

45 Unfortunately we cannot increase the methods in the main text by any significant amount due to the word limit. However, we did make a number of important changes in our wording to make the logic of our analysis clearer and

hopefully understandable by reading only the main text. Readers will still need to refer to the ESM for some details given the word limit.

L23: "Tart cherry in Pennsylvania was not included in this analysis due to insufficient data". Please provide more detailed information (quantitative data) here clarifying why tart cherry could not be included in these analyses.

46. We dropped the data set on tart cherry in Pennsylvania prior to running the analysis because it had the smallest sampling effort of any dataset (see table below). In this decision we also took into account that we had another data set on tart cherry in Michigan which was included in the analysis, as it had almost 4 times more data; in contrast, the smallest data set we retained was our sole data set on almond.

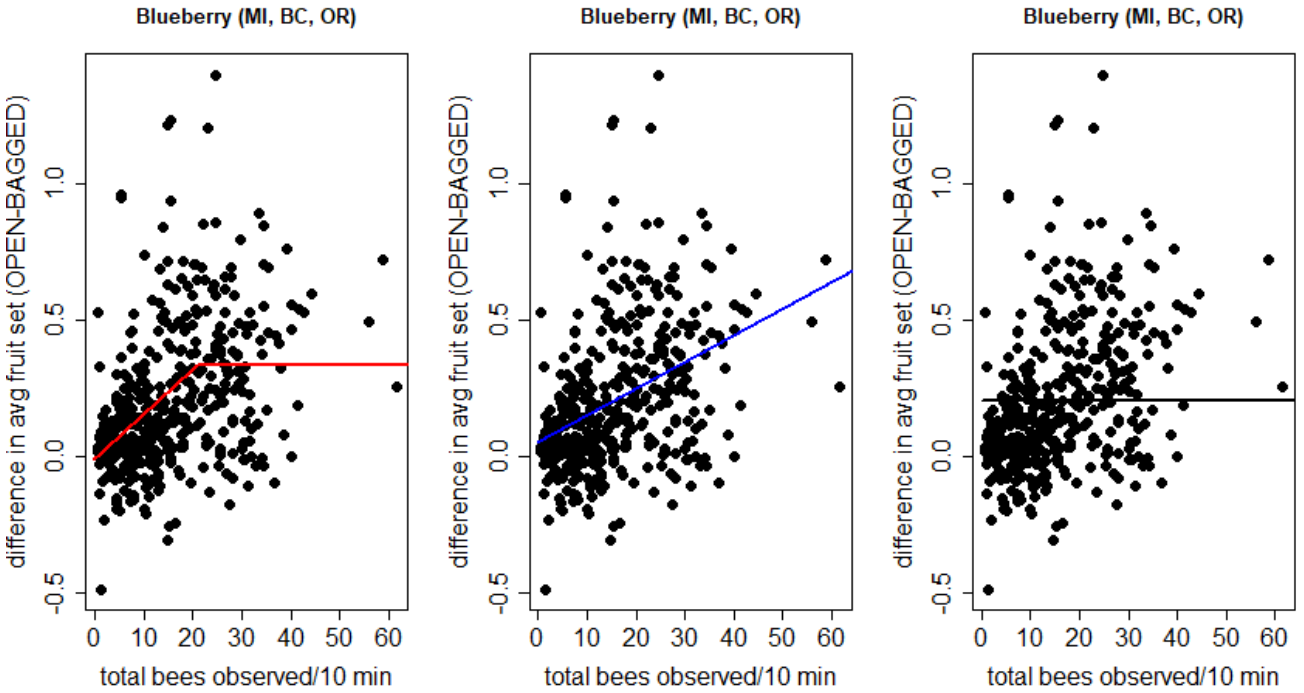
crop	minutes of observation
pumpkin	6615
watermelon FL	5344
watermelon CA	4542
blueberry MI	3760
blueberry BC	3540
sweet cherry WA	1680
blueberry OR	900
tart cherry MI	580
apple MI	570
blueberry FL	420
apple PA	320
almond	235
tart cherry PA	**160**

L187: Supplementary hand pollination (flowers receiving pollen from hand-pollination on top of the pollen they receive by pollinators; flowers not bagged) or hand pollination of bagged flowers?

47 The supplemental hand pollination was done on top of whatever pollination they were getting from bees. Those were open clusters that we added pollen to multiple times during bloom to ensure maximum pollen deposition. We clarified this on lines 236-238 of the ESM.

L191: Why only berry weight? What about fruit set?

48 For our blueberry studies, we used berry weight instead of fruit set because it is known that yields in northern highbush blueberry are more typically increased through larger berry size rather than number of berries, and mean berry weight is often correlated with pollination (Isaacs et al. 2010, Benjamin et al. 2014). If we had used fruit set we would have obtained the same answer, though. See figure below for analysis of relationship between blueberry fruit set and bee visitation. As in the analysis of berry weight presented in the paper, AIC analysis strongly prefers the segmented model here (deltaAIC = 8.6) over the linear model. No relationship model is much worse (deltaAIC = 76).



L198-207: I think it is a strength of this study that you assess pollination limitation also with an alternative method (using a hand pollination approach). Different methods have their particular strengths and potential limitations. For, example, focusing on yield is probably of more direct relevance to growers. On the other hand, estimates of pollination limitation following this approach may be somewhat less precise and more correlational, as many other factors (mentioned by authors) affect crop yield, not only pollination services. The hand pollination approach on the other hand could provide a more direct and potentially more precise method to estimate pollination limitation, but it might be less directly relevant to farmers and final crop yield. I think mentioning the hand pollination results and discussing the advantages and disadvantages of both approaches could be valuable for readers (see also general comment above).

49. We agree that the confirmation using independent methods is important, and we added a discussion of this to the manuscript on lines 250-255. See response #21 above.

L58-62: It would be helpful to included sample sizes in Table S2, especially for the per-visit efficiency data. Table S2 shows that efficiency data were not available for all non-Apis bee species groups, and probably only a fraction of the species in these groups. It would be almost impossible, and probably also not needed to have such data available for all wild bee species of a system. Deriving a single value from a very broad wild bee group including a large number of species differing in traits likely affecting pollination efficiency, however, likely ignores considerable within-group variation in efficiency. I do not think this lack of data invalidates such analyses and outcomes, but I find it important that it becomes very clear on what basis in terms of available data these values were estimated and that the limitations und large uncertainties involved in such an approach should be explicitly discussed in the main text of the manuscript. In particular assuming honeybee efficiencies for many wild bee species/species groups due to a lack of data and using efficiency values estimated for apple in cherry could be problematic and should at least be critically discussed.

50 Sample sizes are now included in Table S2. See response #18 above.

Some efficiency data for Sweet cherry pollination would be available for a European region (Eeraerts, M., Vanderhaegen, R., Smagghe, G., & Meeus, I. (2019). Pollination efficiency and foraging behaviour of honey bees and non□Apis bees to sweet cherry. Agricultural and Forest Entomology).

51. Thanks for pointing us to this new paper. We have updated Table S2 to use this paper's estimates for cherry, and re-run the relevant analyses. See revised results on lines 268-276.

L83-94: This paragraph is identical to such a paragraph in the main text…

52. We have fixed this. See lines 106-109 in ESM.

L115-122: I appreciate this discussion of potential uncertainties and limitations of such extrapolations here.

53. We're glad you found it helpful.

L161-165: In my opinion the strong effect of whether or not to subtract variable costs from crop production values should be mentioned and discussed in the main text (see general comment).

54. See response #41 above. We have added another sentence to clarify.

References cited in the Response to Review

Bartomeus I, Potts SG, Steffan-Dewenter I, Vaissiere BE, Woyciechowski M, Krewenka KM, Tcheulin T, Roberts SPM, Szentgyorgyi H, Westphal C, Bommarco R (2014) Contribution of insect pollinators to crop yield and quality varies with agricultural intensification. PeerJ 2: e328.

Benjamin F, Reilly J, Winfree R (2014) Pollinator body size mediates the scale at which land use drives crop pollination services. J. Appl. Ecol. 51: 440-449.

Garibaldi, L et al. (2013) Wild pollinators enhance fruit set of crops regardless of honey-bee abundance. Science 339, 1608–1611.

Isaacs R, Kirk A (2010) Pollination services provided to small and large highbush blueberry fields by wild and managed bees. J Appl. Ecol. 47: 841-849.

Knight T, et al. (2005) Pollen limitation of plant reproduction: Pattern and process. Annu Rev Ecol Evol Syst 36:467–497.

Park M, Raguso R, Losey J, Danforth B (2016) Per-visit pollinator performance and regional importance of wild *Bombus* and *Andrena* (*Melandrena*) compared to the managed honey bee in New York apple orchards. Apidologie 47 (2) 145-160.

Stavert JR, Liñán-Cembrano G, Beggs JR, Howlett BG, Pattemore DE, Bartomeus I. 2016. Hairiness: the missing link between pollinators and pollination. PeerJ 4:e2779 https://doi.org/10.7717/peerj.2779

Willmer , P G & Finlayson , K 2014 , Big bees do a better job : intraspecific size variation influences pollination effectiveness, Journal of Pollination Ecology , vol. 14 , no. 23 , pp. 244-254

Winfree R, Williams N, Dushoff J, Kremen C (2007) Native bees provide insurance against ongoing honey bee losses. Ecol. Lett. 10, 1105–1113.

Winfree R, Fox J, Williams N, Reilly J, Cariveau D (2015) Abundance of common species, not species richness, drives delivery of a real-world ecosystem service. Ecology Letters 18: 626-635.

Appendix B

Associate Editor Board Member
Comments to Author:
The authors have done an excellent job at incorporating the suggestions made by the reviewers and by myself, which has clarified and improved several important aspects of the manuscript. One of the reviewers who reassessed the manuscript found that the definition of "yield" is still unclear. This reviewer made a couple of additional minor comments. I thus encourage the authors to consider these additional comments and to submit a revised version of the manuscript incorporating these suggestions.

AUTHORS' RESPONSE:
1. Thank you for your helpful suggestions--we also believe the manuscript has been significantly improved. In the new version we have incorporated the reviewer's suggestions for dealing with the yield vs crop production issue and the other points, which we hope will provide clarity.

Reviewer(s)' Comments to Author:

Referee: 1

Comments to the Author(s).
The authors have thoughtfully and appropriately responded to most of the comments made in the first review, and the manuscript is now very much improved.

2. Thank you for your careful reading of the ms and comments.

My primary on-going concern is with regards use of the word 'yield'. In my original review I commented as follows:

Across the different production metrics in this study only two crops (Table s4, pumpkin and melon) were assessed on a per area basis. For most crops in the study the measures were fruit per flower or branch, which are well short of true yield. Getting a true yield measure is difficult in the experimental context, so this is not surprising. Nevertheless it is important to be careful in the choice of words.

To which the authors responded:

13. We agree. Clarification added on lines 161-165.

However, the lines numbered 161-165 do not relate to this problem, so this I an error.
I think the actual words of clarification are at lines 146-148, which has the following text:

Thus our use of the term yield necessarily differs somewhat by crop, but matches commonly used proxies for yield (Garibaldi et al. 2013, Bartomeus et al. 2014).

In my view this text is not sufficient to solve the problem. It remains the case that the word yield appears 55 times in the main document and 18 times in the ESM. Most significantly it appears in the

title and the abstract. The reader has to dig quite deep to establish that for most crops in the study what is actually measured is a proxy for yield, and that no assessment is made regarding how effective the proxy measure is. This matters because the focus of the study is on economic benefits for agriculture, and in agricultural terms yield is understood as a per land area measure. In some circumstances one might examine the yield relative to another high cost input, such as irrigation water or fertilizer. But looking at 'yield' on a per flower or per branch basis is a biology perspective that is poorly linked to any economic measure.

I do not consider this a hard problem to solve. When referring to measures such as fruit per flower or fruit per branch, replace the word yield with the less specific "production". Some uses of the word yield are appropriate because they are made in reference to theory, or to other studies, or to the crops where production was measured on a per area basis (i.e. watermelon, pumpkin). But when referring to the measures such as fruit per flower or fruit per branch, replace the word yield with the less specific "production".

3. We appreciate the reviewer's perspective on this issue and suggestions for how to make our wording more precise. We have decided to change the title to use the term "production" instead of "yield" as suggested, and have also made this change on lines 40, 116, 122, 137, 150, 151, 152, 165, 169, 225, and 631 in the main text and lines 8, 10, 16, 215, 222, 237, 311, 313, 314, 344, 372, 373 (table heading), 393, and 396 (table heading) in the supplement. In a few places, it seemed most appropriate to write "crop production or yield" to reflect the different measurements across crops.

We have also expanded the discussion of the yield vs production terminology on lines 154-158.

Additional minor comments
I suggest inserting the following recent paper (from Proc Roy Soc) which is relevant at line 80
"… , Rader et al. 2016) and that the diversity of wild bee visitors is higher when crops are grown in their biogeographic region of origin (Brown and Cunningham 2019)."
Brown J, Cunningham SA 2019 Global scale drivers of crop visitor diversity and the historical development of agriculture Proceedings of the Royal Society B 286, 20192096

4. Text and reference added on lines 85-86.

Regarding the following response to reviewers :
14. Yes, cultivar and site are confounded. From our perspective, this is not necessarily undesirable…
The explanation given is a fair one, but this thinking is not reflected in the paper. It would be appropriate to mention this confounding in the supplementary methods.

5. We have added an explanation of this issue on lines 26-32 in the supplementary methods.