nature portfolio

Peer Review File

Trade-offs between grassland plant biodiversity and yields are heterogenous across Germany

Corresponding Author: Mr Dario Schulz

Version 0:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Mr Schulz,

Your manuscript titled "Mow it or grow it: heterogeneous biodiversity-yield tradeoffs in grasslands" has now been seen by 3 reviewers, whose comments are appended below. You will see that they find your work of some potential interest. However, they have raised quite substantial concerns that must be addressed. In light of these comments, we cannot accept the manuscript for publication, but would be interested in considering a revised version that fully addresses these serious concerns.

Specifically, for publication in Communications Earth & Environment to be appropriate, we will need you to:

1) Provide compelling new insights into the spatial heterogeneity of the influence of grassland management intensity on yield and species diversity, on the country scale, and discuss the underlying ecological mechanisms

2) Make a robust case that mowing frequency can be used as a proxy for grassland management intensity.

We hope you will find the reviewers' comments useful as you decide how to proceed. Should additional work allow you to address these criticisms, we would be happy to look at a substantially revised manuscript. If you choose to take up this option, please either highlight all changes in the manuscript text file, or provide a list of the changes to the manuscript with your responses to the reviewers.

Please bear in mind that we will be reluctant to approach the reviewers again in the absence of substantial revisions.

If the revision process takes significantly longer than three months, we will be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We are committed to providing a fair and constructive peer-review process. Please do not hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the reviewers' comments with a list of your changes to the manuscript text (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and any completed checklist:

Link Redacted

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

Please do not hesitate to contact us if you have any questions or would like to discuss the required revisions further. Thank you for the opportunity to review your work.

Best regards,

Jialiang Cai

Internal Editor
Communications Earth & Environment

EDITORIAL POLICIES AND FORMAT

If you decide to resubmit your paper, please ensure that your manuscript complies with our editorial policies and complete and upload the checklist below as a Related Manuscript file type with the revised article:

Editorial Policy Policy requirements (Download the link to your computer as a PDF.)

For your information, you can find some guidance regarding format requirements summarized on the following checklist: (https://www.nature.com/documents/commsj-phys-style-formatting-checklist-article.pdf) and formatting guide (https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf).

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

The manuscript aims to evaluate the effect of moving frequency as a proxy of management intensity on the species richness (plant species richness). The authors derived all information from GIS imaging and used large inventory of grasslands over whole Germany. I realised that to the methodological part my knowledge is very limited – so I cannot evaluate these sections of the manuscript. I can comment therefore on the overall message and main findings.

While I read the ms with great interest, I found that its main message is very simplified and it was formerly validated by others – of course not using such a large dataset like this paper – that intensification in grasslands decrease the species richness. This is nowadays rather a triviality. As the authors also stated that the increase in grazing frequency is only a small part of intensification which is also involves the increased use of fertilisers (both mineral fertilisers and dung), sometimes irrigation, and overseeding by productive cultivars. It is also stated that intensification has lower impact in the already intensified and therefore species-poor grasslands than in extensively managed ones with higher species richness. This is also a triviality. In intensively managed grasslands only species with higher disturbance tolerance and broader environmental niche (common grasses and forbs) survive which are tolerate higher magnitudes and frequency of disturbances. Thus, we should not expect that they will vanish with the increase in mowing frequency.

I felt problematic that the occurrence of many grasslands is not driven by macroclimate, and many species rich grasslands occur because edaphic drivers in a particular place, and more species rich than predicted by climatic variables. It was also stated by many authors that species richness itself it is not a good predictor for habitat quality as it is matters which species contribute to the pool – i.e. habitat specific species pool matters not the whole species pool. This cannot be evaluated by GIS imaging. See paper about habitat specific species pools by Helm et al. (2015).

Helm, A., Zobel, M., Moles, A.T., Szava-Kovats, R., Pärtel, M., 2015. Characteristic and derived diversity: implementing the species pool concept to quantify conservation condition of habitats. Diversity and Distributions 21, 711-721.

Based on these above arguments, beside of the intellectual challenge evaluating such amount of GIS analysed data, the paper has not much applied value for research or conservation practice.

Reviewer #2 (Remarks to the Author):

The researchers report on work to correlate numbers of mowing events and changes in the species richness of pastures in Germany. The work is based on error-assessed estimations from remotely sensed images plus ancillary data. The report describes what seems a careful and appropriate analysis. I have a few overarching comments and a few notes.

The authors suggest that mowing is a proxy for grassland use intensity, including grazing behaviors. This is only partially true. I know the authors are aware, but mowing removes biomass whereas grazing includes consumption and respiration, but a portion of the biomass is excreted by the animal in feces and urine, affecting nutrient concentrations. Animals waste part of what they consume, leaving the material on the landscape, trample vegetation, and compact soil. I suggest the authors either downplay the sweeping surrogacy of mowing in their writing or review the differences; the first pathway may be most reasonable.

The authors may address potential bias associated with confusion over cause and effect. Does less frequent mowing increase species richness? Or are areas with fewer more dominant species mowed more often because of greater productivity? Perhaps competition between species is decreasing production, etc.?

Given the central importance of the richness map and the mowing events map to the work, and the nature of the journal having Methods follow Discussion, I suggest that a brief report of the assessed accuracy of those surfaces be included at the

start of the Results. Otherwise it remains a question in the reader's mind throughout. Related to that, I wouldn't leave these important values to a figure in supplemental information (line 336); cite it in the body of the work … how well do these maps perform? The 60% accuracy cited for mowing is a helpful example.

The title indicated in the supplementary materials, "Heterogeneous impacts of mowing frequency on plant species richness and associated yield trade-offs" seems more fitting than the two-part title used, but as the editor suggests.

The language in the manuscript is overly technical in my view, without an attempt to make the work more accessible. I suggest an edit so that the potential readership of the work may be expanded. Frequent use of the term causal does not add clarity for me, especially in a correlative work. This is an explicit comparison in the Discussion (lines 229-230) but this appears correlative work to me, so I am unsure of the use of the term. Examples of text that may be streamlined include: (lines 38-39) "… the causal identification of large scale real world evidence of the management intensity-biodiversity-yield gradient …". May "The core principle behind this approach is to estimate non-observed counterfactual plant species richness under a hypothetical mowing frequency …" (lines 67-68) be made more clear? What does counterfactual mean in this context? What is a "regional production structure"? (line 360) Lines 386-389: "We are interested in the grassland species richness for the counterfactual scenario of a different mowing frequency. We build upon the potential outcomes framework where treatment effects are estimated by comparing observed outcomes to counterfactual outcomes under an alternative treatment." That is tough-going! Might you just say "We assessed changes in grassland species richness under the observed rate and other rates defined in scenarios"? The paragraphs that follow may be simplified as well.

Line 75 requires edit. "This allowed us to inform policy makers so that they may draw conclusions …"

Line 78. The 30 by 30 goal should be cited. It is not common knowledge.

Lines 83-84. Per year? Or over the 4 years?

Lines 144-145. "… our common support assumption holds, namely that the probabilities of different mowing frequencies overlap." I don't know what this means. They overlap in what way?

The sentence at the end of line 158 may be removed.

The English is good but somewhat broken. I suggest minor edits in that regard to the manuscript.

NUTS-3 (line 363) should be cited.

Reviewer #3 (Remarks to the Author):

This manuscript provides an approach to estimate large scale, causal, spatially explicit effects of mowing frequency on plant species richness using remotely sensed data, and assess the tradeoffs between yield and plant species richness. Overall, the research design is novel, and the findings are solid and of interest to a broad spectrum of readers. Moreover, the research provides useful spatially explicit information to balance the agricultural production and biodiversity conservation, which is then important for achievement of SDG goals. However, some major revisions are needed before it can be published.

General comments
1 Grasslands consist of many types whose dominant ecosystem functon(s) differ. For example, the production function for alfalfa fields is dominant, while more species of weeds are not beneficial; rural and urban greenspaces are mainly for entertainment and relaxing, and the biomass from mowing events is often not used for agricultural purpose (no monetary value). Assessing tradeoffs between yield and plant species richness makes sense for management only if these two functions are dominant for that grassland type.

2 The authors examined the spatially-explicit relationships between yield and plant species richness, and identified the drivers of their spatial heterogeneity. As can be seen Fig. 1d, the negative relationships are prevailing across the country, although they vary spatially. But you mentioned that the relationships from the previous experiments are mixed. I suggest the authors discuss the underlying ecological mechanisms of your results more deeply.

3 The authors used mowing frequency as a proxy to grassland management which also includes measures such as grazing regimes, fertilization, and so forth. The authors validated this by making analysis with a composite dataset of grassland management. I suggest the authors examine the relationships between these two datasets (Areas with more mowing events mean more intensified management?) to give more stronger support of such proximation.

Specific comments

Line 19-21: What does the 30 by 30 goal mean? Please rewrite the last sentence of the Abstract to make it more clear and straightforward.
Line 58-59: As you have mentioned above, the grassland management might include mowing, fertilization, and grazing regimes. Are there any relationships between mowing frequency and fertilization, grazing regimes? Please explain.

Line 61-62: The best resolution of Sentinel-2 image is 10 by 10 m. How can you use data at such resolution to estimate plant species richness at finer 16-m2 plots? Please explain.
Line 106-107: The word ‘extensification’ here means the decrease in mowing frequency, right? If so, please clarify it so that it is easy to understand with the cited Fig. 1d.
Line 111-119: From supplemental material, the spatial resolution of these driver datasets are different. Specially, some datasets are available at district level. How did the authors process them to match the parcel scale?
Line 129: For Figure 1d, are there any places where the values are positive? It makes sense to add the break value of zero in classification so that the positive and negative relationships are distinguished clearly.
Line 133: What does “…more marked for very species rich parcels” mean?
Line 157: When calculating opportunity cost, did you consider the factors of labor cost, the investment of fertilizer and mechanics? While the hay value might contain these costs, without deducting them would lead to overestimates of opportunity costs. Please clarify it.
Line 330-331: I am confused about this sentence, “grasslands whose ……”. How does this happen? What is the frequency?

Author Rebuttal letter: The author's response to these comments can be found at the end of this file.

Version 1:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Mr Schulz,

Your manuscript titled "Mow it or grow it: heterogeneous biodiversity-yield tradeoffs in grasslands" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at https://www.nature.com/documents/commsj-file-checklist.pdf .

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit https://www.nature.com/commsenv/open-access

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:
Link Redacted
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Heike Langenberg, PhD
Chief Editor
Communications Earth & Environment

On X(Twitter): @CommsEarth

REVIEWERS' COMMENTS:

Reviewer #2 (Remarks to the Author):

I appreciate the time the authors put to responding to my queries and concerns. The readability of the manuscript is improved. I have some minor comments.

Line 69. What two mowing regimes?
Line 89. Although “we” is used quite frequently in the manuscript, “The authors” does not seem a good alternative. Maintaining an active voice seems best. Perhaps cite the test used and then report the results.
Line 95. This sentence is dense and needs unpacked. Other factors definitely DO affect species richness. This sentence may benefit from editing knowing that readers will not have read methods yet (or ever).
Line 130. “Moving” should be “Mowing” I suspect.
Line 153. I would hesitate to describe a regression explaining 60% of variation as “strong”. Perhaps “moderate”.

Reviewer #3 (Remarks to the Author):

The authors well addressed the questions/suggestions I raised, and I am satisfied with their revised manuscript. Therefore, I recommend accepting the manuscript for publication in current form.

Author Rebuttal letter: The author's response to these comments can be found at the end of this file.

RESPONSE TO REVIEWER COMMENTS:

We would like to thank the editor and reviewers for their valuable feedback. In the following, we have addressed each point raised by the editor and the reviewers (see responses in green). In particular, we clarified our main contribution and our understanding of causality. We explain this in more detail in the responses below as well as in the updated version of the manuscript. Key aspects of the revisions to the initially submitted manuscript include:

- A more comprehensive communication regarding the main contributions of our study: Our main contribution is the quantification of the heterogeneous relationship between mowing frequency and species richness relationship at field level, and its economic valuation.
- A clarification of what is understood by causal mechanisms in the context of this study: Causality implies the causal attribution of a change in outcome (here species richness) to an intervention (here: change in mowing frequency).
- Terminological clarification and discussion regarding species richness/habitat quality/mowing events. More specifically, we point out that our chosen indicators are appropriate for our approach, rather than optimal for a comprehensive in-depth assessment.

In addition, we made small edits to improve the language. A list of references used in the answers here can be found at the end of this document.

Reviewer #1 (Remarks to the Author):

The manuscript aims to evaluate the effect of moving frequency as a proxy of management intensity on the species richness (plant species richness). The authors derived all information from GIS imaging and used large inventory of grasslands over whole Germany. I realised that to the methodological part my knowledge is very limited – so I cannot evaluate these sections of the manuscript. I can comment therefore on the overall message and main findings.

While I read the ms with great interest, I found that its main message is very simplified and it was formerly validated by others – of course not using such a large dataset like this paper – that intensification in grasslands decrease the species richness. This is nowadays rather a triviality. As the authors also stated that the increase in grazing frequency is only a small part of intensification which is also involves the increased use of fertilisers (both mineral fertilisers and dung), sometimes irrigation, and overseeding by productive cultivars. It is also stated that intensification has lower impact in the already intensified and therefore species-poor grasslands than in extensively managed ones with higher species richness. This is also a triviality. In intensively managed grasslands only species with higher disturbance tolerance and broader environmental niche (common grasses and forbs) survive which are tolerate higher magnitudes and frequency of disturbances. Thus, we should not expect that they will vanish with the increase in mowing frequency.

Thank you. We agree that the novelty of our results lies not primarily in providing proof of the relationship. Instead, we quantify the spatial variation in the relationship between mowing frequency and species richness and show that it is very heterogeneous across space within Germany. This is highly relevant to inform the design of agri-environmental payment schemes and related decision support tools for biodiversity monitoring. We also add to the literature by providing large-scale and causal assessments of heterogenous effect sizes using an innovative combination of methods and high-resolution data. We also provide an economic interpretation and quantification of the associated management trade-offs. The actual contribution of our paper is now stated more clearly in

the revised manuscript. The fact that our results are consistent with expectations and previous findings underscores the robustness and validity of our approach, as we relied on independent datasets and purely empirical estimation of these relationships. We rephrase our key findings to better reflect what our (novel) contribution is in the Introduction (l. 39 ff.) and Results section.

I felt problematic that the occurrence of many grasslands is not driven by macroclimate, and many species rich grasslands occur because edaphic drivers in a particular place, and more species rich than predicted by climatic variables.

You correctly point out that our species richness map does not in fact rely on either climatic or edaphic variables, but only on multitemporal, multispectral reflectance measured at 10 m resolution via Sentinel-2, as indicated in lines 65 and 382. We use climatic and edaphic variables only to predict the propensity (likelihood) of a given mowing regime and its subsequent impact on species richness in the following year. Our aim is not to assess habitat quality compared to a hypothetical baseline reflecting pure climatic and edaphic conditions, i.e., before the onset of intensive agriculture (see also next comment). Instead, we set the current de-facto conditions, including observed management as an anthropogenic disturbance, as our baseline and compare it to potential outcomes. In other words, our outcome of interest is not the absolute number of species, but their change - for which we take into account edaphic and climatic drivers. This goal is now formulated more clearly in the introduction (l. 42 f.). Furthermore, we clarify and justify our conceptual choice in the Methods section (l. 463 ff.) and discuss the limitations of our chosen species indicator more thoroughly in the Discussion section (l. 301 ff.).

It was also stated by many authors that species richness itself it is not a good predictor for habitat quality as it is matters which species contribute to the pool – i.e. habitat specific species pool matters not the whole species pool. This cannot be evaluated by GIS imaging. See paper about habitat specific species pools by Helm et al. (2015).

Helm, A., Zobel, M., Moles, A.T., Szava-Kovats, R., Pärtel, M., 2015. Characteristic and derived diversity: implementing the species pool concept to quantify conservation condition of habitats. Diversity and Distributions 21, 711-721.

Thank you for this interesting reference. This comment is directly linked to the previous one.

The provided reference describes a biodiversity index that captures the relationship between habitat-typical species abundance and the abundance of species that emerged in the habitat (invasive or native) as the result of anthropogenic modifications, for example following historical agricultural intensification. The index relies on expert knowledge to determine the „characteristic“ species pool of an ecosystem or region. It provides a tool to monitor habitat quality and biodiversity integrity relative to a historical baseline.

We agree with the reviewer in that our employed plant species richness is not the ideal proxy for biodiversity (if there is any) because it does not capture all relevant qualitative dimensions, including the aforementioned assessment of habitat-specific species composition. Yet, it is a well-suited indicator for this large scale, and our analysis provides important methodological steps that can be extended with other data in future analyses. We now explicitly mention this limitation of our species richness indicator, justify it, and suggest potential ways to overcome some limitations in the discussion section. While we agree that it is desirable to use more accurate indicators such as biodiversity or habitat quality, limited data do not allow for the inclusion of such an indicator. However, upon the availability of such data, our approach can be readily adapted to any other biodiversity metric, including the suggested „Index of Favourable Conservation Status“ – provided that it has been quantified across the sample.

It should be noted that our approach does not rely on remote sensing or GIS per se; rather, we utilize datasets derived with these tools that allowed other researchers the cost-efficiently scale of data collection to national scale. As noted above, our contribution is the quantification of counterfactual species richness under a hypothetical mowing regime. We believe qualitatively superior proxies of biodiversity will become available in the near future, and our approach can easily accommodate such alternative indicators. Therefore, we are demonstrating the application of a process pipeline, rather than suggesting that this is the only and best way to measure biodiversity impacts.

As a potential avenue for conducting such analyses at large scale with the index proposed by Helm et al. (2015) would be to first model the spatial distribution of „characteristic" species based on biophysical & climatic conditions, and second measure the de-facto („derived") presence of single species, e.g. via surveys or remote sensing. The resulting mapped index could then be directly integrated into our proposed workflow. However, this is beyond the scope of the current analysis. We now include the provided reference as an example of a potential indicator for future analyses (l. 311 f.).

Based on these above arguments, beside of the intellectual challenge evaluating such amount of GIS analysed data, the paper has not much applied value for research or conservation practice.

We used your feedback to clarify our contribution (Introduction), and its limitations and possible extensions (Discussion). Thank you. While we relied on remote sensing and GIS tools for data curation and result visualization, our actual analysis employs state-of-the-art econometric methods for causal inference (generalized random forests), rather than GIS. As stated in our discussion, our approach is transferable to other management- and biodiversity indicators and provides spatially explicit information that is valuable for policy targeting. We expand the list of potential use cases of our approach to more clearly illustrate its value for both researchers and conservation practitioners. In addition to our previously described use case of result-based payment schemes, we include: 1) serving as baseline input for monitoring, reporting, and verification of biodiversity credits in sustainable supply chains; 2) supporting regional land use planning (e.g. zonation). We now mention these applications in the discussion (l. 361 ff.).

Reviewer #2 (Remarks to the Author):

The researchers report on work to correlate numbers of mowing events and changes in the species richness of pastures in Germany. The work is based on error-assessed estimations from remotely sensed images plus ancillary data. The report describes what seems a careful and appropriate analysis. I have a few overarching comments and a few notes.

Thank you. We'd like to emphasize that our core contribution is to move beyond correlations. We aimed at establishing causality by employing a double-robust causal inference framework. We leverage machine learning to generate counterfactuals, which are valid under the assumption of conditional unconfoundedness and common support. Both of these assumptions are explained in the methods section (l. 502 ff.) and further discussed in the Supplementary Information (Figure S1 and S9). This is a major contribution of our work, which we now highlight more explicitly in the introduction.

The authors suggest that mowing is a proxy for grassland use intensity, including grazing behaviors. This is only partially true. I know the authors are aware, but mowing removes biomass whereas grazing includes consumption and respiration, but a portion of the biomass is excreted by the animal

in feces and urine, affecting nutrient concentrations. Animals waste part of what they consume, leaving the material on the landscape, trample vegetation, and compact soil. I suggest the authors either downplay the sweeping surrogacy of mowing in their writing or review the differences; the first pathway may be most reasonable.

Thank you. In the revised manuscript, we exercise greater caution in presenting our findings. In the title and conclusion, we exclusively refer to mowing frequency rather than grassland management intensity. We point towards its potential as a proxy for grassland management intensity in the abstract and introduction, while clearly outlining its limitations in the discussion.

Specifically, we strengthen our argument that mowing frequency is a suitable proxy for grassland management intensity in the following ways (l. 402 ff.). First, we provide more details on the mowing data we used, characterizing and discussing the advantages and limitations put forward by the authors of this paper. Second, we add references on grassland management intensity indicators to justify the use of mowing frequency as a proxy for management intensity (l. 404). Third, using spatially coinciding observations in our dataset, we show that the estimated impact from the alternative management intensity indicator provided by Lange et al. (2022) is of comparable magnitude (when looking at the same year and standardized, see Table S3), and also correlates strongly (r= 0.778), as now indicated in the updated version of the manuscript (l. 154).

Finally, we also point out that future research shall take advantage of even further developed remote sensing products that more accurately and reliably reflect grassland management intensity. Our approach can easily accommodate such new products and integrate them at scale.

The authors may address potential bias associated with confusion over cause and effect. Does less frequent mowing increase species richness? Or are areas with fewer more dominant species mowed more often because of greater productivity? Perhaps competition between species is decreasing production, etc.?

Thank you. We revised to make this relationship clearer. Our approach aims to establish a causal impact assessment of mowing on species richness by a) leveraging time, i.e. the impact of mowing on species richness in the subsequent year rather than the same year (l. 65 ff.); b) controlling for a wide range of observed and latent (unobserved) environmental and production-related characteristics that affect both mowing and species richness (see methods section on contextual variables). The reason why we believe to also control for a set of unobserved characteristics is explained in the directed acyclical graph shown in Supplementary Figure S9 and indicated in l. 487 ff. Specifically, a key advantage of generalized random forests (compared to linear regression), is their ability to leverage non-linear combinations of all features for prediction, thereby representing to some extent also non-observed features.

We fully agree that species richness and yields are endogenously related and refrain from making any causal claims with respect to yields. Instead, we already referred to associations, not causes throughout the manuscript, for example in the introduction (l. 47 ff.), results (e.g., l. 178 and Title of Figure 2 Panel C), and Supplementary Note S2 on yield calculation. We make this now more explicit throughout the manuscript (l. 181 ff.).

Given the central importance of the richness map and the mowing events map to the work, and the nature of the journal having Methods follow Discussion, I suggest that a brief report of the assessed accuracy of those surfaces be included at the start of the Results. Otherwise it remains a question in the reader's mind throughout. Related to that, I wouldn't leave these important values to a figure in supplemental information (line 336); cite it in the body of the work … how well do these maps perform? The 60% accuracy cited for mowing is a helpful example.

We now mention the product accuracies in the results section. Specifically, we mention the estimated overall accuracy in cross validation as reported in the original papers, namely 60% for the mowing frequency and R^2=0.42 for the plant species richness. In addition, we performed our own cross validation with independent (although not fully comparable) data on species richness as described in the SI Supplementary Note S1 and Supplementary Figure S8 and indicated in the manuscript in lines 286 ff.

The title indicated in the supplementary materials, "Heterogeneous impacts of mowing frequency on plant species richness and associated yield trade-offs" seems more fitting than the two-part title used, but as the editor suggests.

Apparently, we forgot to update the title in the SI. Although we prefer the current title (Mow it or grow it), but we will accept the editor's suggestion here.

The language in the manuscript is overly technical in my view, without an attempt to make the work more accessible. I suggest an edit so that the potential readership of the work may be expanded.

Thank you for pointing this out, We add additional several clarifications regarding some of the technical vocabulary. Specifically, we made the following edits on technical, yet unavoidable jargon:

- We clarify endogeneity in our context as "mowing affect yield, but expected yield also affect the mowing decision" (l. 182)
- We clarify counterfactual and potential outcomes as an approach that "tries to answer the question: what species richness would we observe on each specific field if the field had been mowed differently?" (l. 474 ff.), and distinguish it from the similar, yet different concept of "dark diversity", which may be more known among ecologists (l.469)
- We exemplify the "common support assumption" and "propensity score", see lines 158 ff.

Frequent use of the term causal does not add clarity for me, especially in a correlative work. This is an explicit comparison in the Discussion (lines 229-230) but this appears correlative work to me, so I am unsure of the use of the term.

We now clarify what we mean by "causal" in the beginning of the results section "causal effect of mowing frequency on plant species richness" (l. 95 f.) by saying "This relationship is causal in the sense that changes in mowing frequency (but not other factors) influence plant species richness (but not vice-versa)." Thus, we isolate the effect of mowing frequency from all other factors that determine plant species richness.

Furthermore, we refer to non-causal associations between mowing frequency and yields throughout the manuscript, namely in the introduction (l. 47 ff.), results (e.g., l. 178 f. and Title of Figure 2 Panel C), and Supplementary Note S2 on yield calculation. We clarify this non-causal link explicitly in the manuscript (l. 181 ff.).

Examples of text that may be streamlined include: (lines 38-39) "… the causal identification of large scale real world evidence of the management intensity-biodiversity-yield gradient …".

Suggested new sentence: „Existing studies frequently rely on experimental evidence, yet they lack causal identification of large-scale real-world evidence regarding the management intensity-biodiversity-yield gradient and quantification of the underlying heterogeneity." (l. 36)

May “The core principle behind this approach is to estimate non-observed counterfactual plant species richness under a hypothetical mowing frequency …” (lines 67-68) be made more clear? What does counterfactual mean in this context?

- In this sentence, the term "counterfactual" refers to something that is not currently observed, but rather imagined or hypothesized. Specifically, it refers to the plant species richness that would exist under a hypothetical mowing frequency that has not been observed or implemented. This term is often used in causal impact analysis to compare observed outcomes with what would have happened under different conditions, allowing researchers to assess the effects of interventions. In this context, the "counterfactual plant species richness" represents the richness of plant species that would exist if the mowing frequency were different from what is currently observed. (l. 68 ff.)
- We rephrased the section; it now reads: “The core principle behind this approach is to estimate non-observed plant species richness under a different mowing frequency than the observed one. Since it is impossible to observe and compare these two mowing regimes, we use a comparison group based on similar mowing propensities, but different observed mowing regime. These mowing propensities are calculated from observed variables and their non-linear interactions and used to weight observations for valid inference.”
- We also clarify counterfactual and potential outcomes as an approach that “tries to answer the question: what species richness would we observe on each specific field if the field had been mowed differently?” (l. 474 ff.), and distinguish it from the similar, yet different concept of “dark diversity”, which may be more known among ecologists (l.468)

What is a “regional production structure”? (line 360)

Regional production structure refers to variables taken from the agricultural census, available for the year 2016 at municipal level. We took the values from 2016 to avoid reverse causality, since changes in field-level grassland management intensity could change overall structural characteristics. We thereby rule out potential feedback loops that would bias our causal estimates. Specifically, these variables are: average farm size, percentage of organic farms, percentage of permanently cultivated area, percentage of permanent grassland, percentage of cattle farms out of all farms, and cattle density per hectare. These variables were selected because they explain variation in grassland management intensity, plant species richness and yields in the long run, since related structural changes occur rather slowly compared to field-level management decisions and weather. We now clearly explain this in the manuscript (l.429 ff.)

Lines 386-389: “We are interested in the grassland species richness for the counterfactual scenario of a different mowing frequency. We build upon the potential outcomes framework where treatment effects are estimated by comparing observed outcomes to counterfactual outcomes under an alternative treatment.” That is tough-going! Might you just say “We assessed changes in grassland species richness under the observed rate and other rates defined in scenarios”? The paragraphs that follow may be simplified as well.

We add a clarification to make it more accessible to a wider audience maintaining our claim of causality rather than correlation based on the described methods employed. We add alternative statements/formulations to our results section to ensure readers from diverse backgrounds can relate to what we mean (at the risk of being slightly repetitive). Specifically, we clarify counterfactual and potential outcomes as an approach that “tries to answer the question: what species richness would we observe on each specific field if the field had been mowed differently?” (l. 474 ff.), and distinguish it from the similar, yet different concept of “dark diversity”, which may be more known among ecologists (l. 468 f.)

Line 75 requires edit. “This allowed us to inform policy makers so that they may draw conclusions …”

We corrected the grammatical error (changes underlined); the new revised sentence reads: „This allows us to inform policy makers so that they may draw conclusions for the effective targeting of conservation policies accounting also for provisioning services and thus food security.

Line 78. The 30 by 30 goal should be cited. It is not common knowledge.

We add a citation along with a description of the goal. The revised sentence is: „More specifically, we estimate opportunity cost of marginally extensifying 30 percent of German grasslands to address the 30 by 30 goal to protect 30% or terrestrial ecosystems by 2030, as specified in the new Global Biodiversity Framework (CBD, 2022)“. The citation refers to: Parties to the United Nations Convention on Biological Diversity. Kunming-Montreal Global Biodiversity Framework. CBD/COP/15/L25 (2022).

Lines 83-84. Per year? Or over the 4 years?

Per year; we specify this in the new sentence: „The observed annual mowing frequency ranges between 0 and 5.5 with a mean of 1.9 (SD=0.82).“ We also added an explanation that a continuous mowing frequency is the result of our field -wise spatial aggregation of pixels.

Lines 144-145. “… our common support assumption holds, namely that the probabilities of different mowing frequencies overlap.” I don’t know what this means. They overlap in what way?

Our propensity score represents the likelihood of an individual field receiving a particular treatment (i.e., mowing frequency) based on their observed characteristics (predictor variables). It serves as a summary measure of the factors influencing mowing frequency. The common support assumption posits that there exists overlap in the propensity scores between fields that were de-facto mown with different frequencies, indicating that there are observations in both groups with similar probabilities of being mown with a given frequency. For example, there are observations with a predicted mowing frequency of 2 (predicted by our random forest propensity model, not the original mowing events map), but a de-facto (i.e., observed) mowing frequency of 3. These observations build the comparison group against which observations with a de-facto (observed) mowing frequency of 3 will be compared to get the causal impact of a change in mowing frequency. In the context of our analysis, the "common support assumption" refers to the premise that the probabilities associated with different mowing frequencies overlap. This implies that there is a shared range where the probabilities of observing each mowing frequency intersect. The propensity score quantifies the comparability between treated and untreated groups by capturing their underlying characteristics, while the common support assumption ensures that this comparability holds true across the entire range of propensity scores.

We add this explanation to the manuscript for improved clarity (l. 158 ff.)

The sentence at the end of line 158 may be removed.

We removed the sentence „To this end, we combine three steps.“

The English is good but somewhat broken. I suggest minor edits in that regard to the manuscript.

We checked for grammar mistakes, sentence simplifications and made numerous small edits throughout the manuscript.

NUTS-3 (line 363) should be cited.

We reformulate so that the new sentence reads (addition underlined): "Therefore, we include socioeconomic variables relating to the structure of the agricultural economy such as the average farm sizes or share of organic farms at district level (i.e., NUTS-3, Nomenclature of Territorial Units for Statistics, 3rd level) in the year 2016". In addition, we add a reference to the European Comission for more details on the NUTS in the Supplementary Information.

Reviewer #3 (Remarks to the Author):

This manuscript provides an approach to estimate large scale, causal, spatially explicit effects of mowing frequency on plant species richness using remotely sensed data, and assess the tradeoffs between yield and plant species richness. Overall, the research design is novel, and the findings are solid and of interest to a broad spectrum of readers. Moreover, the research provides useful spatially explicit information to balance the agricultural production and biodiversity conservation, which is then important for achievement of SDG goals. However, some major revisions are needed before it can be published.

General comments

1 Grasslands consist of many types whose dominant ecosystem functon(s) differ. For example, the production function for alfalfa fields is dominant, while more species of weeds are not beneficial; rural and urban greenspaces are mainly for entertainment and relaxing, and the biomass from mowing events is often not used for agricultural purpose (no monetary value). Assessing tradeoffs between yield and plant species richness makes sense for management only if these two functions are dominant for that grassland type.

This is a valid point. Thank you. We excluded urban green spaces from the analysis by applying an agricultural grassland mask and contain at least one pixel that was mown at least once to ensure fields are comparable with respect to their land use and primary purpose of agricultural production. Within these agricultural fields, we assume the main purpose to be agricultural production. Furthermore, we restrict our analysis to permanent and cultivated grasslands as identified by the product derived by Tetteh et al. (2024). We now state this more clearly in the manuscript (l. 450 ff.) and explicitly restrict the validity of our findings to this subset of grasslands. We also refer more clearly to permanent agricultural grasslands in the abstract.

2 The authors examined the spatially-explicit relationships between yield and plant species richness, and identified the drivers of their spatial heterogeneity. As can be seen Fig. 1d, the negative relationships are prevailing across the country, although they vary spatially. But you mentioned that the relationships from the previous experiments are mixed. I suggest the authors discuss the underlying ecological mechanisms of your results more deeply.

The evidence we are aware of points towards a negative relationship between number of cuts and species richness, but results are heterogeneous (not mixed). We corrected this statement in the introduction (l. 34).

We contextualize our results by suggesting that the smaller effects in the south and northwest are likely due to the lower prevailing plant species richness in these intensively used regions, resulting in a smaller impact of extensification on this outcome and possibly partly due to environmental factors that favor higher biomass production and mowing frequencies, such as less sandy soils and more precipitation. We also extend our discussion to ecological mechanisms that can explain the heterogeneity in the mowing-species richness relationship across Germany. We explain our results by considering the intermediate disturbance hypothesis, the timing of first mowing, and the role of mowing and grazing (l. 137 and lines 256 – 285). The specific ecological mechanisms are linked the species functional traits, and – while highly relevant – considering trait-specific effects is beyond the scope of this analysis.

3 The authors used mowing frequency as a proxy to grassland management which also includes measures such as grazing regimes, fertilization, and so forth. The authors validated this by making analysis with a composite dataset of grassland management. I suggest the authors examine the relationships between these two datasets (Areas with more mowing events mean more intensified management?) to give more stronger support of such proximation.

Thank you for this suggestion. We checked the correlation between mowing frequency (by Schwieder et al. 2022) and the grassland management intensity index (by Lange et al. 2022). They correlate moderately (r = 0.25). Nevertheless, we then show that the estimated impact from the alternative management intensity indicator is of similar magnitude (when looking at the same year and standardized, see Table S3), and correlates strongly (r= 0.778) with our main estimates, as now indicated in line 153 f.

Specific comments

Line 19-21: What does the 30 by 30 goal mean? Please rewrite the last sentence of the Abstract to make it more clear and straightforward.

We clarify the sentence in the abstract (addition underlined): „Motivated by the 30 by 30 goal to protect 30 percent of terrestrial ecosystems by 2030, we estimate opportunity cost of marginally extensifying 30 percent of German grasslands.“ We also added a reference the first time we mention the goal in the introduction (line 79 ff.), as well as in the results section (line 216 ff.).

Line 58-59: As you have mentioned above, the grassland management might include mowing, fertilization, and grazing regimes. Are there any relationships between mowing frequency and fertilization, grazing regimes? Please explain.

All three dimensions contribute to grassland management intensity, as previously suggested by Blüthgen et al. (2012), and indicated in the Methods section (l. 412 f.). According to these authors, these dimensions affect to nutrient loadings (in particular C/N/P) in soil and biomass, as well as the

habitat suitability for species. In general, fertilization and mowing frequency are positively correlated, because fertilization increases the nutrient availability in the soil, which can stimulate grass growth and lead to higher biomass accumulation. In turn, higher biomass may necessitate more frequent mowing to maintain the desired vegetation height and prevent overgrowth. In contrast, the grazing regime can be either positively linked to mowing frequency, for example when mowing complements rotational grazing to promote regrowth, but it can also be negatively associated because intensive grazing may eliminate the need to control vegetation growth my mowing. We added the following clarification:

"Apart from mowing frequency, the intensity of grassland management is commonly characterized based on fertilization input and grazing pressure because these dimensions also affect nutrient loadings in the soil and flora. Fertilization and mowing frequency are often positively correlated because higher nutrient availability stimulates biomass growth and thus enables or requires more frequent mowing. Grazing can be both positively or negatively associated with mowing depending on whether it complements or supplements each other in the management strategy." (l. 412 ff.)

Line 61-62: The best resolution of Sentinel-2 image is 10 by 10 m. How can you use data at such resolution to estimate plant species richness at finer 16-m2 plots? Please explain.

This comment touches two important points:

1) Detectability from 10m resolution: At first glance, it may not sound intuitive that 10m sensors are able to detect the presence and count single plant species, that are often just a few centimeters in size. However, as explained in detail in the original paper by Muro et al. (2022), the temporal variation in the spectral signature especially during the greening and browning phase enables the detection of single functional groups. The prediction we use here does not refer to specific plants, but to an aggregate number of how many different functional plant (groups) can be expected within a given area.
2) Transferability across resolutions: We recognize that the difference in sensor resolution and ground-truth plot sizes may confuse readers and try to clarify the divergence, and its implications for our research along with assumptions we made more clear in the method section, in addition to the explanation already contained in the Supplementary Information. It's important to note that altering the units from the resolution of the satellite imagery (10 by 10 meters) to match the desired plot size (16 square meters) isn't straightforward. Such a conversion would necessitate knowledge of the species-area relationship, which can vary significantly across ecosystems and be non-linear in nature. In other words, the expected number of species to be found increases with plot size, but one cannot expect to find twice as many species (on average) on a plot of twice the size. Attempting this conversion therefore introduces additional uncertainties without necessarily adding value to the analysis. Our research operates under the solid assumption, as documented in previous studies such as Fischer et al. (2010), that the 4 by 4 m area where ground truth species data are collected is representative of the coarser resolution plot from which the spectral signals are gathered. This assumption serves as a fundamental basis for our methodology and interpretation of results. Thus, while it may seem intuitive to match the resolution of the satellite imagery to the desired plot size, doing so overlooks the complexities of ecological dynamics and the inherent limitations of remote sensing data. Ultimately, our approach prioritizes robustness and consistency in analysis over attempting to match resolutions that may not align with the ecological reality of the study area.

Line 106-107: The word 'extensification' here means the decrease in mowing frequency, right? If so, please clarify it so that it is easy to understand with the cited Fig. 1d.

We clarify this and split the sentence for better readability. It now reads: „Panel D shows the geographic distribution of extensification impacts, i.e., the impact of a unit reduction in mowing frequency. The largest gains from management extensification in terms of plant species richness occur in the northeast, while the lowest gains occur in the more intensively managed northwest and south of Germany."

Line 111-119: From supplemental material, the spatial resolution of these driver datasets are different. Specially, some datasets are available at district level. How did the authors process them to match the parcel scale?

Indeed, the datasets differ in spatial and temporal resolution. The majority of data sources was available in raster format with resolutions ranging from 10-1000 meters, as indicated in Table S1. Only administrative indicators, namely protected areas and district level agricultural census variables were obtained as polygon shapes. Our analysis is based on field shapes, available as polygons. We extract the raster data within each polygon by taking the area weighted mean of all pixels within the polygon. We state this more clearly in the method section (l. 453 ff.).

Line 129: For Figure 1d, are there any places where the values are positive? It makes sense to add the break value of zero in classification so that the positive and negative relationships are distinguished clearly.

The share of positive values is 0.34%; this is not visually representable. We describe this in the Figure note. (l. 147 f.)

Line 133: What does "…more marked for very species rich parcels" mean?

We reformulated the sentence in the Figure note to "In contrast to that, the impact of mowing frequency becomes drastically stronger for very species rich fields". We add the following interpretation in the main text: „The top 5% species-richest fields exhibit a three fold higher partial effect of mowing frequency compared to 75% of fields with lower species richness".

Line 157: When calculating opportunity cost, did you consider the factors of labor cost, the investment of fertilizer and mechanics? While the hay value might contain these costs, without deducting them would lead to overestimates of opportunity costs. Please clarify it.

Thank you for this important point. We acknowledge the possible relevance of these costs in agricultural operations (and other costs, e.g. machinery etc) may matter too. These costs are highly case specific and depend on many (farm and farmer specific) factors we cannot observe. Considering the uncertainties involved, we preferred to use upper bound estimates of opportunity costs, as stated in line 175 f. „We aim to derive spatially explicit upper bound opportunity costs for field owners in three steps". Our reasons are as follows:

Our decision was primarily influenced by the lack of spatial variability in the data on these costs. From a practical perspective, calculating the exact opportunity costs of a unit change in mowing frequency requires many assumptions, for example, on the production system (feed own livestock, silage, dry and sell hay on market…), mechanization level, fuel prices, transportation costs, fertilizer quantities depending on soil conditions and management intensity, among others. Without accurate data on how these costs vary across the study area, subtracting them from the analysis would introduce additional uncertainty without providing substantial value to the results.

Furthermore, if costs items were constant across space we would subtract a fixed value from the numerator of our cost-effectiveness estimate. While this adjustment would impact the absolute values

of the opportunity costs, it would not alter their relative positions to each other. Thus, it would not affect the spatial pattern displayed on the map or the overall scale of our abatement cost curve.

Given the substantial uncertainty involved and the marginal value gained from this adjustment, we concluded that the potential benefit does not outweigh the added complexity. We now indicate more clearly throughout the manuscript that we mean upper bound opportunity costs.

Line 330-331: I am confused about this sentence, “grasslands whose ……”. How does this happen? What is the frequency?

We clarify the sentence: „Thereby, we excluded grassland fields, for which the spectral-temporal signature differs from the training fields “ The frequency of exclusion across all grasslands reported in secondary data is stated in the following sentence: „Based on the area of applicability, the prediction is valid for 70% of grassland areas.“ However, as described in our sampling strategy, we excluded fields smaller than 0.3 hectare which are most prone to measurement errors. For this reduced sample of 1,669,838 fields , we obtain 84% overlap with our species prediction data within the area of applicability (i.e., the pixels that have a similar feature space to the training pixels). We only mention the more conservative overlap estimate of 70% in the manuscript for simplicity.

References

1. Blüthgen, N. et al. A quantitative index of land-use intensity in grasslands: Integrating mowing, grazing and fertilization. Basic and Applied Ecology 13, 207–220 (2012).
2. Fischer, M. et al. Implementing large-scale and long-term functional biodiversity research: The Biodiversity Exploratories. Basic and Applied Ecology 11, 473–485 (2010).
3. Griffiths, P., Nendel, C., Pickert, J. & Hostert, P. Towards national-scale characterization of grassland use intensity from integrated Sentinel-2 and Landsat time series. Remote Sensing of Environment 238, 111124 (2020).
4. Lange, M., Feilhauer, H., Kühn, I. & Doktor, D. Mapping land-use intensity of grasslands in Germany with machine learning and Sentinel-2 time series. Remote Sensing of Environment 277, 112888 (2022).
5. Muro, J. et al. Predicting plant biomass and species richness in temperate grasslands across regions, time, and land management with remote sensing and deep learning. Remote Sensing of Environment 282, 113262 (2022).
6. Schwieder, M. et al. Mapping grassland mowing events across Germany based on combined Sentinel-2 and Landsat 8 time series. Remote Sensing of Environment 269, 112795 (2022).
7. Tetteh, G. O., Schwieder, M., Blickensdörfer, L., Gocht, A. & Erasmi, S. Agricultural land use (vector) : National-scale crop type maps for Germany from combined time series of Sentinel-1, Sentinel-2 and Landsat data (2017 to 2021). https://doi.org/10.5281/ZENODO.10619782 (2024).

Response to reviewer comments (second revision)

In the following, we address all reviewers' comments in green.

Reviewer #2 (Remarks to the Author):

I appreciate the time the authors put to responding to my queries and concerns. The readability of the manuscript is improved. I have some minor comments.

Thank you for your time and effort. We addressed all of your comments as described below and feel that the manuscript has gained clarity.

Line 69. What two mowing regimes?

Thank you for pointing out this unclarity; we specify it to "a higher and a lower frequency mowing regime".

Line 89. Although "we" is used quite frequently in the manuscript, "The authors" does not seem a good alternative. Maintaining an active voice seems best. Perhaps cite the test used and then report the results.

We now explicitly state that we are referring to "The authors of the underlying mowing event dataset".

Line 95. This sentence is dense and needs unpacked. Other factors definitely DO affect species richness. This sentence may benefit from editing knowing that readers will not have read methods yet (or ever).

Thank you for highlighting this potential misunderstanding. We changed the sentence to: "This relationship is causal in the sense that we isolate the causal effect of mowing frequency from other observed and unobserved factors that explain both mowing frequency and plant species richness. We mitigate bias from reverse causality by looking at plant species richness in the subsequent year."

Line 130. "Moving" should be "Mowing" I suspect.

Thank you for identifying this typo; we corrected it in this line (130) and one other line (91).

Line 153. I would hesitate to describe a regression explaining 60% of variation as "strong". Perhaps "moderate".

We changed the word to "moderate", which surely is more accurate and in line with conventions for the wide target audience.

Reviewer #3 (Remarks to the Author):

The authors well addressed the questions/suggestions I raised, and I am satisfied with their revised manuscript. Therefore, I recommend accepting the manuscript for publication in current form.

Thank you for your time and effort. We appreciate your previous comments and recommendation for publication.