Grasslands make up most agricultural land and provide fodder for livestock, yet data on grassland yields is limited due to direct use on farms. Accurate yield information is crucial for informing policy and understanding ecosystem services and inter-annual variations. In this study, we estimate annual grassland yields in the Ammer catchment area of southern Germany for 2019 using three approaches: (i) a model combining field samples, satellite data, and mowing information (RS), (ii) the biogeochemical process-based model LandscapeDNDC (LDNDC), and (iii) a rule-based approach based on field measurements and spatial productivity data (RVA). All approaches yield similar results, estimating yields of 4-9 t/ha, with some spatial variations. LDNDC generally produces higher yields, especially for the first cut and grasslands mown one or two times per year. Mowing frequency was the most significant factor influencing yields, with no major differences in the impact of abiotic factors (e.g., climate or elevation) across approaches. This comparison offers new insights into the strengths and limitations of each method and highlights the importance of grassland productivity maps for long-term studies on climate and management impacts.

This study is a pioneering effort to compare entirely different approaches for estimating grassland biomass and provides valuable insights for future land use estimations. I recommend it for publication, with minor to moderate revisions focusing on enhancing clarity and broadening its relevance for a wider audience.

We would like to thank you for taking the time and reviewing our manuscript. We appreciate the positive feedback and comments. We are confident that the revisions made in the manuscript based on the suggestions of the reviewer will enhance its quality. We answer to each comment below in italics while the reviewer's comments are upright.

General comments:
#1 The key findings highlight the spatial and temporal differences between the three approaches. However, these differences are difficult to interpret in terms of the underlying reasons why each model or approach yields distinct results, and it is unclear which approach performs better overall. The results section presents only the spatial and temporal disparities without delving into potential explanations. Factors such as climate forcing uncertainties in the LDNDC model input, the quality of data, methods used for remote sensing photo analysis, and limitations of survey data could all influence the outcomes.

To enhance the clarity and depth of the analysis, I recommend the following:

1. A more comprehensive comparison, either in a table or in the discussion section, outlining the pros and cons of each method.
2. An in-depth analysis that investigates the potential root causes behind the observed differences in results, including an exploration of the most significant factors affecting each approach (beyond mowing).
3. A discussion on the future implementation of these approaches, specifying under which conditions each model is more suitable than the others.

Thank you for these suggestions. Regarding your first point, we agree that a table would improve the clarity and show pros and cons in a concise manner. Such a table would be rather qualitative as the limited number of in-situ data doesn't allow for a quantitative in-depth assessment, unfortunately. We would still like to follow this idea and add a table with pros and cons to the revised version of the manuscript, most probably in the conclusions.

Regarding your second point, we think that such an in-depth analysis of influencing factors would be very interesting. We analyzed the influence of the most relevant factors for grassland yield, namely mowing frequency, temperature, precipitation and elevation (compare chapter 4.3) by assessing the correlation between these factors and the estimated yields and the plots. We hope that the results are formulated clearly but will revise the manuscript to make sure that our findings are well understandable. We suggest that we could add additional plots as Figure 9 and statistics between influencing factors and yields per mowing frequency (for all or only the most dominant mowing frequencies). In that regard the influence of temperature, precipitation and elevation is investigated apart from the mowing frequency. However, we checked the results and they are very similar to the Figure 9.

Regarding your third point, we agree that the discussion so far lacks specific suggestions for future implications. We will add this to the discussion as far as it is possible.

#2 Each method, due to its underlying model mechanism and input data, introduces uncertainties. However, the study presents only the estimated annual AGB from the three approaches without providing an uncertainty range for each estimation. To draw a more reliable conclusion about the similarity of the three results, it is essential to include upper and lower bounds for each estimate. I recommend incorporating the potential uncertainties associated with the input data for each method and providing confidence intervals (upper and lower bounds) for the annual AGB estimations.

Thank you for this remark. We agree that it is important to include uncertainty information to modelled results. For the RS and LDNDC approaches the validations result in r2 and RMSE (root mean square error) estimates, which are described in L340 and L345. To analyze the sensitivity of the models towards the input data would indeed give a more sophisticated insight into model uncertainties. However, there are no uncertainties related to all input datasets, hindering us from applying an error propagation analysis. In addition, this seems to be out of scope for this manuscript. We hope that providing the estimated errors derived from the validation based on in-situ measurements is sufficient. We will add error bars to the annual yield estimates of Figure 3 to highlight these.

#3 The results suggest that all approaches are highly sensitive to mowing dates, with the estimated annual grassland yields showing a strong correlation to the number of mowing events. To assess the stability and effectiveness of each model, I recommend conducting a sensitivity test by removing the number of mowing events from all three models. This would allow for an evaluation of how each model performs with the remaining input variables

Thank you for highlighting this aspect and the suggestion to test this. We agree that the idea to test how the models perform without mowing data sounds interesting as this seems to be the most important input variable. However, all approaches rely on the mowing dates as input. There are no yield estimates without harvesting information. Therefore, we can't test this using these approaches and our model setups, unfortunately. However, as mentioned above we could add additional plots to the Appendix showing the relationships of yields to temperature, precipitation and elevation per mowing frequency.

#4 The annual grassland biomass estimation is based on a particular region in this study. It would be better to give some insights in the discussion section to show the possibility and potential use case to apply those three approaches to a bigger region, what the limitations (such as data availability) would be, and what method(s) will be most likely to serve better.

Thanks for this important comment. In the current version, we missed discussing the transferability of the approaches to other regions. We will add this to the discussion in the revised manuscript.

Specific comments:

#1 Figure 3. Related to my comment #2, try to add an uncertainty range for the annual AGB

Thank you for this suggestion. We will add the estimated errors to the annual yields of Figure 3 to show uncertainty ranges.

#2 Figure 5. Try to reverse the color bar by using red as positive and blue as negative
Thanks for this comment. We will follow your suggestion and change the colors to make the plot more intuitive.

#3 Line 430 - 440 & Figure 9. Is the correlation between annual yields and temperature, precipitation, and elevation obtained without dropping dominant mowing events? If so consider eliminating mowing events from the model and test the sensitivity

Thank you for your remark. The mowing events were not dropped here. But we tested the relationships per mowing frequency (not shown in current manuscript version yet). The results stayed relatively the same. Nevertheless, as suggested above, we could add the Figures and statistics per mowing frequency to show that the relationships stay constant among varying mowing frequencies.