nature portfolio

Peer Review File

Plant diversity within communities, not among them, stabilizes grassland productivity across spatial scales

Corresponding Author: Professor Xiang Liu

This file contains all reviewer reports in order by version, followed by all author rebuttals in order by version.

Version 0:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Major comments
This study is exceptional in the spatial extent and high replication of the underlying data, covering a large area of China. The collection of plant trait data and soil chemistry over such a geographic extent, I believe specifically for this study, is a remarkable achievement. Finally, this study is notable for its integration of field data and remote sensed NDVI data, and its use of piecewise structural equation models (pSEMs) to understand drivers of diversity and stability. I found the conclusions, including the association of fast strategy plants with stability, interesting and well supported by the pSEM. These conclusions will be a useful contribution to the nascent field of spatial scaling of biodiversity-stability relationships. I reviewed the deposited R code and found it well commented and comprehensive.

1. I do have some concerns with the way alpha stability was calculated, the use of data dredging and backwards selection approaches, and the many different analyses applied to the same variable set. I detail these concerns below ("Methodological comments").

2. The writing was generally smooth, but I found a few inconsistencies in the arguments presented, detailed below ("Writing comments").

3. In some cases, the authors did not find support for hypothesized relationships between functional diversity and stability, or between beta richness and spatial asynchrony. Here the authors could have usefully considered the possibility that the hypotheses were sound but applied either to traits not measured or to larger distances between the subplots. No rationale was given for the spatial distances between the subplots (relatively small, the centers of 30 x 30 m subplots were 50m apart – so the gap between subplots was presumably 20m): was there any empirical data suggesting that metacommunity processes like plant dispersal were important in the range 20-50m for example?

Methodological comments

4. Methods: “Alpha stability of subplots was defined as the weighted average of community stability across subplots.” Please explain what you mean by weighted? Following your formulae, it looks like you summed the NDVI-based productivity value across all four subplots, then summed the standard deviation of NDVI productivity across all four subplots, and then took the ratio of these values. However, I think you should have instead calculated the mean/sd value (inverse cv) for each subplot separately and then averaged these four values. Alpha stability is the denominator of spatial asynchrony so this calculation will affect this metric too. I agree with the calculation of gamma stability.

5. Statistics. I am puzzled why the authors include individual linear model analyses (gls) in addition to the piecewise SEM. The pSEM is superior to all individual linear model approaches as it explicitly allows for a causal structure, whereas the linear models run the risk of including spurious variables or excluding indirect effects. Further, it literally is data dredging to use the dredge function in MuMin, where all possible models are run, relative importance of each variable is assessed and model selection is used to choose the most efficient model. Statisticians specifically counsel against data dredging approaches.

6. The a priori SEM is a series of hypotheses about how the system works. I disagree with removing non-significant paths in pSEMs to improved the model fit (as judged by Fisher's C?) as this would be altering your a priori hypotheses. If a path is not significant in a pSEM, it may still be affecting the overall model, for example via indirect paths and hidden covariance. By imposing a backwards selection approach on their pSEMs, the authors forego their hypothesis-based approach. Backwards selection is another approach generally discouraged by statisticians. The total effect of an explanatory variable on a response variable, including indirect effects, can be calculated by the semEff package which uses bootstrapping to assess significance; this may be a better way to summarize your model.

Writing comments.

7. At first reading, these two sentences in Abstract seem contradictory: "In contrast, we found no evidence that variation in species richness and functional diversity among local communities contribute to stabilize productivity at larger spatial scales. Although the effects of environmental conditions on local diversity and stability differed between the two regions, the relationships between diversity and stability remained consistent at both the local and larger spatial scales." After reading the manuscript I understand the results that you are trying to convey, but the abstract should be understandable at first reading.

8. Introduction. Line 54: "First, resource acquisition strategies determine species' responses to environmental changes, with fast species exhibiting faster recovery but lower resistance and slow species having higher resistance but slower recovery. In this context, while communities dominated by fast species can exhibit low community stability due to low resistance, communities dominated by slow species can exhibit high community stability due to high resistance ". This seems contradictory: First sentence establishes fast species have resilience and slow have resistance, yet second sentence says that only slow have stability as they are resistant (but resilience is also a type of stability). I think here the issue is that the literature perhaps emphasizes slow strategies and resistance, but in your Discussion you propose that fast strategies lead to resilience that also stabilizes productivity. The Introduction is perhaps caught between these two ideas, but you need to either set up two alternate hypotheses at the start (one resistance based, one resilience based) OR you need to restrict the Introduction to describing only the more common hypothesis.

9. Line 60: "species can either respond asynchronously to environmental fluctuations when they differ in their functional traits (Fig. 1a2), or display competition-induced negative covariance if they have similar traits (Fig. 1a3). Therefore, communities with low or high functional diversity in fast-slow traits are likely to show greater stability due to compensatory dynamics or negative covariance" This seems contradictory: Greater stability due to compensatory dynamics or negative covariance is, according to the first sentence, only possible under high diversity, not low diversity (contradicts second sentence).

10. Figure 1 needs much more explanation in the legend, describing what the different colours represent, what all the abbreviations mean, what level the traits are calculated at (species? community?) in each part of (a), and how part (a) and (b) relate to each other. It is also not stated what the trait is, as there are many traits where a high CWM (community weighted mean) trait value would not be expected to lead to high alpha stability. The authors are making their argument explicitly for the fast-slow trait continuum, with CWM of the "slow" strategy argued in the Introduction to lead to resistance. Even having read the paper, I am struggling to understand the arguments presented in this figure, suggesting that it should also be better explained in the main text.

11. Statements such as "a positive trend between CWM and alpha stability" are non-informative without knowing what traits loaded positively on the underlying PCA axis. I suggest modifying CWM throughout the manuscript to clarify the direction, e.g. "fast strategy CWM".

12. The logic behind this statement (Line 225) is unclear. Why would the effect of diversity on stability increase with mean diversity? "In our study, Qinghai-Tibet Plateau has locally higher species richness (mean = 12, se = 0.67; Fig. S7a) than Inner Mongolia Plateau (mean = 9, se = 0.36) or the regions analyzed in Ren et al.42 (mean = 8), making alpha richness a more important predictor of alpha stability than the environment."

Minor comments

Line 96 change "along" to "among".

(Remarks on code availability)
Yes I reviewed the R code. The code was linearly arranged and well commented. The link between code and figures was particularly well done. The readme provided a description of all variables in the data set (i.e. meta-data) and information on executing the code.

Reviewer #2

(Remarks to the Author)
SUMMARY

This study examines the relationships among species and functional diversity, and the stability of biomass productivity within plant communities at local (alpha-level) and larger (beta-level) spatial scales. Using data collected in 237 grassland sites in the Qinghai-Tibet Plateau (93 sites) and the Inner Mongolia Plateau (151 sites), the authors assess more specifically the relationships between alpha- and beta-species richness, alpha- and beta-functional composition and diversity, and alpha-biomass productivity stability and spatial asynchrony in biomass productivity. Using Structural Equation Modelling (SEM), the authors further tested the direct and indirect effects of environmental factors (i.e., temperature, precipitation, and soil fertility) on the stability and spatial asynchrony of biomass productivity through their impacts on species richness and the functional composition of plant communities. The results show that species richness and the presence of fast-growing species positively influence the stability of biomass productivity at the local level. In contrast, no relationship was found between beta-richness, beta-functional diversity, and the spatial synchrony of biomass productivity. These findings hold true in both regions; however, the authors further demonstrate that the pathways through which environmental factors influence alpha-biomass productivity and spatial asynchrony in biomass productivity differ between the two regions.

While this study relies on an impressive dataset, I have major concerns regarding this manuscript:

1. Clarity of terminology: The manuscript is very hard to follow due to imprecise use of terminology. For example, the authors use the term "community stability" throughout the Introduction. To me, "community stability" typically refers to the stability of species richness or species abundance (see, e.g., Jarzyna et al. 2022; Ghosh et al. 2024). However, I only realized at the end of the Introduction that "community stability" in this study actually refers to the stability of ecosystem functioning, as captured by plant biomass productivity. Similarly, in Fig. 1, the authors hypothesize that communities with high community-weighted mean (CWM) and low alpha functional diversity, with low CWM but high alpha functional diversity, or with both low CWM and low alpha functional diversity can lead to high alpha stability. However, it is unclear which trait is considered here. For instance, the results would likely differ depending on whether the trait is SLA or LDMC. Other examples of ambiguous terminology are provided in the specific suggestions below. Overall, I believe the manuscript requires clearer definitions and justifications for the metrics used. These metrics should be explicitly defined in the last paragraph of the Introduction, with references to the Methods section for further details. Additionally, I suggest clearly defining the metrics and abbreviations in Figure 1 to enhance readability and understanding.

2. Justification of Structural Equation Modelling (SEM): The authors used Structural Equation Modelling and provided a conceptual figure (Fig. 1) to illustrate the relationships to be tested. While these analyses are well-conducted, I believe the structure of the model should be better justified. I found the manuscript lacking sufficient explanation regarding the links included in the conceptual framework. This information could perhaps be added to the Methods section, included in the legend of Figure 1, or summarized in a table that could be placed in the Supplementary Materials. From my point of view, it is crucial to demonstrate that the SEM structure is based on solid knowledge and strong hypotheses (see Grace 2008).

3. Scale mismatch in data collection: I also have concerns about the mismatch between the scale at which data on plant diversity were collected (i.e., in 4 subplots of 0.5 m × 0.5 m) and the scale at which data on biomass productivity were collected (i.e., 4 subplots of 30 m × 30 m). Although these subplots overlap, such a scale mismatch in data collection might explain why beta-diversity and beta-functional diversity are not influencing the spatial asynchrony of biomass productivity. Since the plant diversity data were collected at a relatively small spatial scale and not upscaled to larger scales, I also suggest toning down the strong conclusions regarding diversity-stability relationships at large spatial scales.

4. Clarity of statistical analyses: Finally, it was not always clear to me, and it was not well explained in the Methods section, why the different statistical analyses were conducted. For example, I do not understand the rationale for using both an ANCOVA and a model selection analysis to test the effects of the different variables on alpha stability and spatial asynchrony. How do these analyses complement each other, and do they not yield very similar results? Providing a clearer justification for the use of these methods and explaining their distinct contributions to the analysis would significantly enhance the clarity and robustness of the manuscript.

Below, I present more specific suggestions that may help, I hope, to improve the manuscript.

SPECIFIC SUGGESTIONS

•L39-41: This section is unclear. What is meant by "environmental change"? Please clarify or provide examples to help the reader understand.
•L51: Does "community stability" here refer to the stability of species richness over time, or do you mean the stability of biomass production? To me, "community stability" typically refers to the stability of abundance or diversity within a community community (see e.g. Jarzyna et al. 2022; Ghosh et al. 2024), rather than the stability of ecosystem functioning or biomass production.
•L54-59: It is hard to see how the mechanisms discussed are related to or reflected in the conceptual framework supporting the SEM presented in Fig. 1. Could you provide a clearer explanation or better link these ideas?
•L72-75: Apologies for pointing this out, but this section is also unclear. Does "temporal stability" here refer to the stability of biomass production, species richness, or something else? Please specify.
•L81: I suggest replacing "essential" with "a strong driver of" for greater precision.
•L98: Once again, the term "community stability" is ambiguous. Please clarify its meaning in this context.
•L101-103: Why do you expect the pathways through which environmental factors affect stability to differ between the two regions? Providing a stronger rationale or referencing relevant studies would help strengthen this point.
•L107: It only becomes clear here that your focus is on the stability of productivity. Consider clarifying this earlier in the manuscript to improve the reader's understanding.

•L121-130: Why are you using both an ANCOVA and a model selection analysis? How do these analyses complement each other? Clarifying their distinct purposes and contributions would be helpful.
•L140-143: This sentence is unclear. In the SEM presented in Fig. 1, there is no direct link between spatial asynchrony and alpha stability, so beta diversity cannot affect alpha stability via spatial asynchrony. Is that correct? Please revise for clarity.
•L156-157: Is it shown somewhere that higher mean annual precipitation and soil fertility contributed to higher spatial asynchrony and higher gamma stability? If so, please provide a reference or elaborate.
•L171-172: You did not assess ecosystem stability at large spatial scales; your analysis was limited to the alpha level for stability if I understood correctly. I think that this conclusion is not supported by your results and appears to be beyond the scope of the study.
•L187: This finding is not related to CWM but rather to the 'slow-fast' PCA axis.
•L204-206: How was this classification performed? I could not find this information in the Methods section. Please provide details.
•L211-212: This contradicts previous studies, such as García-Palacios et al. (2018). Consider discussing how your findings align with or diverge from these studies.
•L268-269: This conclusion is only valid at the local scale, correct? If so, please specify to avoid overgeneralization.
•L285: Did you account for the sampling year in your statistical analyses? If so, please specify how.
•L294: How did you record aboveground biomass? Adding more details would enhance the Methods section.
•L295: Replace "collected" with "measured" for precision.
•L295-296: Did you analyze the Qinghai-Tibet Plateau and Inner Mongolia Plateau sites together? Are the species pools similar in these two regions?
•L316: Specify what "temporal stability" refers to in this context.
•L341-342: Did you calculate the mean and variability of annual temperature and precipitation per year or per growing season?
•L351: Ref. 12 does not seem appropriate here. Did you mean Ref. 56?
•L360-361: This is unclear—weighted by what? Please clarify to avoid ambiguity.
•L394: Why did you use an ANCOVA? What was it used to test? How does it complement the model selection analysis? Providing a clearer rationale would improve the reader's understanding of your analytical approach.
•L406: It would be interesting to examine the response of alpha mean productivity, not just stability. Including this analysis could provide additional insights to your study.
•L419-431: Why was gamma stability not included in the SEM as a function of alpha stability and spatial asynchrony? This omission is puzzling, given its relevance in studies such as Wilcox et al. (2017), Hautier et al. (2020), Qiao et al. (2022).
•Fig. 1: This figure is difficult to read, and some key information is missing. For example, please define terms like PC1, t0, t1, comm1, and comm2 in the caption. Additionally, the phrase "high community-weighted mean" is ambiguous. A high CWM SLA is not equivalent to a high CWM LDMC, and these measures cannot be interpreted in the same way. Specify which CWM is being referred to here.
•Fig. S9: The labels on the x- and y-axes are not readable. Please adjust the font size or formatting to ensure they are clear and legible.

(Remarks on code availability)
The analyses and results of the paper are reproducible and the code can be easily used by the community. A 'README' file is provided and it contains instructions in order to reproduce some of the analyses, figures and results. I however noticed in the code that in some analyses, the variables were log-transformed. This could be detailed in the Methods' section. Note also that all the code is not available, for example I couldn't find the code used to run the model selection analysis and the SEM.

Reviewer #5

(Remarks to the Author)
GENERAL COMMENTS

This manuscript assesses the role of plant taxonomic and functional diversity estimated from field sampling on the stability of productivity (estimated from remote sensing data) in multiple plots located in two different regions of different climates and biodiversity. The analyses take place at two spatial scales (local and regional), comparing large grassland-populated áreas with different environmental conditions.

The work and results fit the scope and relevance of Nature Communications. They are relevant for the ecological community and can be interesting for the remote sensing communities, where there is a growing interest in biodiversity and its relationship with ecosystem functions. Understanding the role of different facets of plant diversity on the stability of ecosystem functions is relevant for assessing the impacts of biodiversity loss, decision-making, and resilience management of natural and anthropic ecosystems. While not the first ones, the work uses a combination of ecological surveys and remote sensing data; the second assesses ecosystem function stability, which requires a long time series of continuous data unavailable from field surveys.

While the presented results support the conclusions, I have some concerns regarding the suitability of the remote sensing metric used as a proxy of ecosystem productivity. This choice might explain some of the results that disagree with those of former studies. New enhanced and more robust remote sensing proxies of ecosystem productivity might cast more reliable results. In particular, the use of remote sensing NDVI as a proxy may be outdated, particularly when looking at short-term

temporal scales. Recent advances in remote sensing, such as the NIRvP index, would offer more accurate proxies of productivity and its stability. NIRvP combines an enhanced indicator of vegetation state (particularly in disturbed and semi-arid sites) with information about the incoming radiation, which might differ across sites. Some of the results presented regarding spatial asynchrony that contradict previous findings might result, in fact, from the use of NDVI, which is sensitive to bare soil and unable to account for across-site differences in incoming radiation.

Also, maybe an important question to be at least clarified is “management”. While it is mentioned as a relevant element in the abstract (once) and the introduction sections (once), the authors do not seem to consider the management regime in the sampling sites of this study. Since management can affect stability, are the sampling plots under different management regimes, and if so, are these known or unknown? If unknown, could these have had any significant impact on the results of this study? If known, how have these been accounted for?

The description of the methods could also be improved, and some questions that could potentially affect the results were clarified. Finally, notice that Nature Communications is multidisciplinary; some concepts should be more clearly presented to the readers, which could not be deeply specialized in the topics of the article, despite their interest.

SPECIFIC COMMENTS

Introduction
Line 62: maybe clarify what you mean by “competition-induced negative covariance”.

Lines 63-65: As phrased, this statement is confusing. Is it trying to summarize both mechanisms or only the second? Please summarize more clearly which conditions lead to each situation (stability or instability)

Lines 72-75: Since alpha and gamma stability are defined as a sort of coefficient of variation of NDVI, how can then their

Line 99: Further specify what “larger spatial scales” are.

Figure 1: Fig 1 is confusing and should be improved.
1) First, the figure caption does not define all the terms included in the figure. What do “t1” and “t2” and the histograms associated represent? Different communities? Different species of the same community? Different traits? Different times? Please clarify.
2) Fig 1a and b represent the trait distributions but do not show the corresponding productivity stability or asynchrony; this would help the least specialized reader to understand what the authors mean by those terms. Also, the slow-fast spectrum is not represented or made explicit in the plots. Should the reader understand that this spectrum is presented in the trait value axis?
3) Fig a1,3 do not seem to represent different CWM. As “(a1) high community weighed mean (CWM) and low alpha functional divergence”; however, the mean of the trait value seems to match the one of Fig. a3, defined as “both low CWM and low alpha RaoQ”. Shouldn’t the distributions of Fig. a1 be shifted towards the right side of the trait value axis? Fig. a1 only seems to present higher frequencies than Fig. a3, which, once normalized, do not modify the CWM.

Methods
Lines 285-286: The sample size of both regions is quite imbalanced; one is close to double the other. Could this induce any bias in the estimation of the regional diversities or variances? Have the authors checked that this does not modify the results, for example, by random resampling of equal sample sizes?

Line 295: I’m surprised that sampling “five healthy, mature leaves 296 from up to five randomly-chosen individuals per species and site” covered almost 90 % of the sampled biomass. For plots with a few species and/or plants with numerous small leaves, five individuals might represent a small fraction of the sample’s weight. How many species were there per plot? Fig. S9 is hard to read, and the presentation of absolute quantities makes it hard to understand whether, for example, plots with low taxonomic richness were systematically undersampled to measure their functional traits. I recommend making Fig. S9 clearer, which also includes absolute values of species richness and sampled biomass, and that the authors consider whether there were systematic biases in the characterization of plant traits.

Line 297: “randomly-chosen” does not need to be hyphenated

Lines 297-299: Did the authors take any measure to prevent water loss (and therefore leaf area reduction) between the sampling and the scanning? (e.g., keeping the samples in hermetic cold environments). If not, how could the area reduction by water loss have affected the results?

Lines 296-307: The fate of the biomass sampled is unclear. Do the authors determine N and P from the very same leaves used to determine SLA? From a subsample of it? When describing SLA analyses, why do the authors reference the amount of biomass, and then, when describing N and P analyses, why do they refer to the fraction of species sampled? Did the authors take two different subsamples from each sample to determine SLA and nutrient contents? If so, is there any systematic difference between the subsamples taken for each trait? What effect could it have on the analyses? Please clarify.

Lines 296-307: The authors used NDVI as a proxy for vegetation productivity, however. While this index has traditionally

been used, recent studies have reported better proxies, in particular, when vegetation cover is low, such as the near-infrared reflectance of vegetation (NIRv ~ NDIV * R_NIR, Badgley, et al. 2017), and even better when the incoming radiation is taken into account (Dechant et al., Dechant, et al. 2022). Have the authors considered using better estimates of vegetation productivity, such as NIRvP? While NIRv can only approach vegetation productivity at monthly/seasonal scales, Dechant et al. show that for shorter scales, the role of incoming radiation is relevant. On the one hand, these estimates could improve the characterization of the stability across sites featuring different illumination conditions; on the other hand, this work could also provide insight into how much ecological studies' conclusions are biased by using NDVI. The near-infrared reflectance (R_NIR) could be extracted from GEE, and incoming PAR could be approximated with the FLDAS Surface downward shortwave radiation (SWdown_f_tavg * 0.45).

Lines 351-365: The use of the different weights is unclear. The authors use "unweighted relative abundance" to calculate Simpson diversity but "weighted" averages of community stability across subplots. How was relative abundance calculated using the number of individuals, the leaf area, the biomass, other…? What are the weights used in the second case, and why do weights differ between metrics? Please clarify.

Lines 375-377: How were the community means weighted, using the number of individuals, the leaf area, the biomass, other…? Please clarify. The same applies for the "unweighted relative abundance" in line 378. Specify if this is the same used for Simpson, and if calculated differently, indicate also how it was calculated.

Line 399: Check the "or", should it be "on", or "or between"? The statement is unclear.

Line 402: "(Fig. 13)", I think the authors mean Fig. S14.

Line 406-409: If the predicted variables (alpha stability and spatial asynchrony) were added to Fig. S14, the figure would also give an idea of how predictors correlate with them.

Results
Figure 3: Figures should be self-explanatory. Please add to the caption what "richness", "Regions", and "richness:Regions" stand for, and as in Figure 5 caption, define the symbology associated with the degree of significance. Do the same for Figure 4.

Lines 114-118: Could the fast-slow functional composition explain the spatial asynchrony? Has this been tested?

Line 112: Define "iv" the first time it is used. It is currently defined in line 414

Discussion
Lines 191-193: Does the referenced study also focus on grasslands? Specify.

Lines 197-212, 221-223: Regarding the recommendation of using NIRv (see methods), notice that NDVI can induce uncertainties in the estimation of productivity when there is bare soil in the pixel. NIRv (and other indices) are more robust to that. Understanding whether NDVI is a good enough proxy to assess the effect of disturbance/aridity on productivity would be interesting.

Lines 234-242: This might result from not accounting for incoming radiation when estimating productivity and its stability. NIRvP-based stability estimates might provide different results (and maybe more consistent with the literature).

Line 270: This statement needs to be corrected: "linked to alpha stability at local, alpha spatial scales"

(Remarks on code availability)
I did not run the code and R is not the programming language I know the best. However, the code seems sufficiently commented and includes human-readable .csv files, a README.txt file describing the other files content and role. The necessary packages are installed when R code is executed.

Version 1:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thank you for this opportunity to review the revisions to this manuscript. I have reviewed all the authors changes, and I am pleased that most of my earlier concerns have been addressed. I appreciated the caveats that the authors added to the discussion regarding spatial scale and unmeasured traits. I still think Figure 1 needs work to reach its potential, and provide some potential solutions below.

Abstract: please change “contributes to (de)stabilize” to “contributes to (de)stabilizing” throughout, to be grammatically correct.

Results: In “…and unrelated to the dominance of fast-growing species, as
indicated by CWM PC1.” I suggest using “summarized” rather than “indicated”

Discussion: I had trouble understanding in this sentence if you were highlighting a difference between the SEM plots and the partial regression plots, or something else: “Although higher
mean annual precipitation and temperature 360 favored dominance by fast-growing species,
this did not translate into higher alpha stability; a pattern that contrasts with results from partial regression plots.”

Figure 1. The figure is still not understandable without a very detailed reading of the legend (which while much more useful, has now become very complicated). Part of the issue with the figure itself is that the SEM uses terms (e.g. CWM PC1, apha RaoQ, beta RaoQ) that are not used in the panels a-d so it very difficult for the reader to relate the different panels of the figure. Why not put on the “trait value” axis the words “slow” and “fast” on the left and right extremes of the axis to indicate the slow to fast strategy trait gradient and modify “trait value” to be “trait value (PC1)”? It seems like the major difference between (a)&(c) versus (b)&(d), is whether functional divergence as measured by alpha RaoQ is low or high, and whether functional diversity among communities as measured by beta RaoQ is low or high. I note that you have carefully drawn each row of panels so the same two communities are shown, but different aspects of these communities are highlighted in different panels on the same row. Consider therefore adding vertical blue panel strips (similar to the horizontal blue panel strips) that say “Low alpha RaoQ, high beta RaoQ ” and “High alpha RaoQ, low beta RaoQ”. If instead you prefer to label (a),(b), (c), (d) individually then labels such as “low alpha RaoQ”, “high alpha RaoQ”, “high beta RaoQ”, “low beta RaoQ” could be used, but then you really don’t need the second community in plots (a) and (c) because this is really a temporal comparison of a community to itself.

Figure 3. Please, in the legend, tell the reader where they can learn the identity of “others” (e.g. a supplementary table, etc). I am also concerned that the authors are showing regression slopes for non-significant relationships. The convention is that regression slopes are only shown when they are significant, otherwise just the cloud of points should be shown.

Figure 4. Please explain the cut-off for significance of paths, as there seems to be a discrepancy in the legend “Non-significant paths are not shown ($P \geq 0.1$). Numbers on arrows are standardized path coefficients (scaled by their mean and standard deviation), and asterisks indicate statistical significance (***$p < 0.001$; **$p < 0.01$; *$p < 0.05$).” My guess is that the authors originally showed marginally non-significant paths ($0.1>p>0.05$), but now do not and forget to adjust the first sentence.

(Remarks on code availability)
Addressed in first review, still acceptable.

Reviewer #2

(Remarks to the Author)
Review
The authors have well addressed the comments raised in the previous review. In line with the recommendations, they have now (i) strengthened the theoretical foundation of the framework they are testing, (ii) clarified their methods and statistical analyses, and (iii) streamlined their analytical approach (for instance, by removing the model selection analysis). Overall, I think that these revisions have considerably improved the manuscript. However, I believe that several additional changes would further enhance the clarity and impact of the paper. More specifically:
1. In the discussion, the authors could further elaborate on the underlying mechanisms and processes at play. In particular, they could revisit the components of stability (resilience and resistance), which are now mentioned in the introduction, and link these mechanisms more explicitly to their results to provide a stronger interpretation.
2. The authors should ensure that the new sections added to the discussion remain fully coherent with the rest of the manuscript. I recommend a careful re-reading to strengthen the logical connections between ideas and to make the arguments more explicit.

Specific comments (line numbers refer to the revised version with tracked changes)
• L34: Why was the number of plots changed in this revised manuscript?
• L40: Should this not read “relationships between environmental conditions and local diversity and stability”?
• L43–46: The sentence “Our study offers new insights into how functional traits along the fast–slow leaf economics spectrum mediate the effects of environmental change factors (e.g., precipitation) on ecosystem stability at contrasting spatial scales” feels somewhat disconnected from the preceding part of the abstract. If you wish to highlight the fast–slow spectrum here, you should also mention results on the relationship between these traits and stability in the abstract I think.
• L109: You also examine the relationships between species richness, functional diversity, and composition, since you consider the effects of CWM as well. Please clarify.
• L126: “CWM PC1” – this variable should be introduced more clearly. As results appear before the methods, a brief explanation here of what is exactly this variable would help the reader.
• L196: From a semantic perspective, I would say that it is preferable to avoid saying that an effect “disappears”. Instead, write that the effect “was not observed”?
• L212–214: Please expand here on the underlying mechanisms. For instance, describe how fast strategies may enhance stability and link back to the mechanisms outlined in the introduction (e.g., improved resilience).

• L294: Would “potentially greater” not be more accurate here?
• L302: Typo: should read patterns.
• L310: Replace “disappeared” with “was not observed”
• L314–320: Would you really expect that considering belowground traits would result in a stronger effect of beta FD on stability? While the discussion of dispersal traits is interesting in this context, I wonder if the section on belowground traits is appropriate here when discussing beta stability.
• L321–334: This section is unclear. I would consider moving this information to the Methods section instead. I understand it was added in response to reviewer comments, but it seems more appropriate for the Methods. Again, ensure that the additions to the discussion remain coherent with the overall flow of the paper.
• L329: The notion of “mismatches” is unclear. I could follow because I had read the review responses, but in the manuscript itself this is difficult to understand. Could you please clarify that section? Again, should it not be presented in the Methods section instead?
• L360: Replace “dominance by fast-growing species” with “the dominance of fast-growing species.”
• L375: Please expand on the implications of your results. For instance, in the conclusion you write: “In conclusion, our results demonstrate that the functional traits of dominant species influence ecosystem stability. While species richness and functional diversity contribute to alpha stability at local scales, they show no relationship with spatial asynchrony at larger scales. The influence of environmental conditions on diversity and alpha stability differs between regions, yet the relationship between alpha diversity and stability is universally consistent.” This is an important conclusion, but its broader ecological or management implications could be made clearer.
• L406: Is the 10.2% referring to overlap in species composition or in the species pool? Please specify.
• L774: Add “positively” before “related” for precision.
• L798–801: Please clarify that the paths from abiotic factors to alpha stability and spatial asynchrony are not shown but are nevertheless included in the SEM.
• Figure 1: Why not include a covariance path between CWM PC1 and Alpha RaoQ? Based on expectations, plant communities dominated by fast-growing species should be less functionally diverse no?
• Figure 1e: You could use a more self-explanatory label for “CWM PC1”
• Figure 4: In panels (a) and (b), only significant paths are shown, whereas in panel (c) both significant and non-significant paths are displayed. Why this inconsistency? It would be clearer to adopt the same approach across panels.

(Remarks on code availability)
The analyses and results of the paper are reproducible and the code can be easily used by the community. A 'README' file is provided and it contains instructions in order to reproduce some of the analyses, figures and results.

Reviewer #5

(Remarks to the Author)
GENERAL COMMENTS
The authors have addressed the questions and suggestions made in the first review. Following other reviewers’ recommendations, SEM analyses have been simplified, and additional checks have been applied. I thank the authors for testing the use of NIRvP; the comparison is valuable for future analyses, even if eventually this index did not yield improved results. The updated version of the manuscript reads more clearly after key statements were rewritten or corrected.

I just have a few suggestions to clarify some aspects that remain confusing

SPECIFIC COMMENTS (lines refer to the version with tracked changes)
Line 31-33: From the statement, it seems that the reason why it is unknown that the relationships hold at larger spatial scales is management; however, the study avoids managed areas. It rather seems that the reason behind this is an open question is it has not been addressed, as later suggested in the introduction. Maybe management is not so relevant in this study as to include it in the abstract.

Lines 265-267: “suggesting that species in the desert steppe tend to aggregate in patches with relative low plant cover due to strong abiotic heterogeneity.” It is unclear where this conclusion comes from. Is it taken from the reference 47 (Sasaki, et al. 2024) or is it made by the authors? This statement requires further explanation and references that support it or a clearer link to the work of Sasaki.

Lines 302-304: If homogeneity can explain the lack of relationship between beta diversity and stability at larger scales, could the steppe sites with high abiotic heterogeneity be used to test whether this hypothesis holds?

Line 362: Point to the figure subplot that the statement refers to.

Line 404: The samples were “dried at 65 °C to a constant weight”; how was this achieved? Were the samples regularly measured or just dried for a sufficiently long time? If so, what was the drying time?

Lines 406-424: The discrepancy between the percentage of biomass that the species used for SLA (89.8%) and NP determination (81.5%) remains unclear. The authors indicate in the answer to the reviewer comments that not all the SLA samples could be used for NP determination due to insufficient sample quantity. I guess this means that some species were not included in the NP analyses, likely due to small leaf sizes.

However, the differences in the shape of the histograms in Fig. S5 make me wonder whether the specie/s not sampled for NP were present in a large number of plots. Also, once I understood that SLA and NP come from the same leaves but that NP analyses were not performed in some cases due to insufficient samples, I wonder how the original Fig. S9 shows a percentage of species biomass lower for SLA than for NP in some cases. I understand some data might have been filtered after determining SLA due to errors or large uncertainty.
How were samples where SLA or NP values were missing handled to compute plant functional diversity? Was there any sort of gap-filling? This all should be clarified in the methods, or if too detailed, in a supplementary section.

(Remarks on code availability)
As in the first review, I did not run the code, as I cannot provide expert advice on programming language I know the best. However, the code seems sufficiently commented and includes human-readable .csv files, a README.txt file describing the other files content and role. The necessary packages are installed when R code is executed and the libraries referenced in the manuscript are included.

Version 2:

Reviewer comments:

Reviewer #1

(Remarks to the Author)
Thank you for these revisions. They are all fine, expect I believe that the authors may have made a mistake in the lettering in the legend of Figure 1. The legend describes the effects of alpha RaoQ on alpha stability as occurring in (a) and (b), but my understanding is that on the actual figure these occur in (a) and (c). Similarly the effects of beta RaoQ on asynchrony are described in the legend as being in (c) and (d), but I see these in the figure itself on panel (b) and (d). My apologies if I've misunderstood - but in that case the legend really is not clear.

(Remarks on code availability)
Code was reviewed on first instance and hasn't appreciably changed.

Reviewer #2

(Remarks to the Author)
General comments

Although the authors have addressed some of the previous revisions, I would still strongly recommend to improve the logical flow of ideas in the discussion and to make the arguments more explicit. As it stands, I found that it remains somewhat superficial. The implications of the findings and the mechanisms involved could be examined in greater depth I think. I would recommend a revision to strengthen this section.

Specific comments (line numbers refer to the revised version with tracked changes)

L31: Consider to add “however” before “whether”.
L117: Is this “a principal axis of the principal component analysis of the CWM of fast strategy traits”, or rather an axis associated with these CWMs ? Please be more precise.
L117: Replace “divergence” with “diversity” for consistency with the rest of the manuscript, and define it explicitly as measured by RaoQ.
L202: Replace “disappeared” with “was not observed” (as requested in the previous review).
L254–260: Please incorporate elements from your response to clarify the reasoning. As currently written, it remains unclear why and how belowground traits would lead to a stronger relationship between beta functional diversity and stability.
Reference to the authors’ response:
“Indeed, we expect that belowground traits may lead to a stronger relationship between beta functional diversity and stability… This response diversity across heterogeneous environments can enhance spatial asynchrony among communities, thereby strengthening the relationship…”
Please integrate this explanation more directly into the revised text to make your argument explicit.
L286: Replace “disappeared” with “was not observed” (as mentioned previously).

(Remarks on code availability)
The analyses and results of the paper are reproducible and the code can be easily used by the community. A 'README' file is provided and it contains instructions in order to reproduce some of the analyses, figures and results.

Reviewer #5

(Remarks to the Author)
GENERAL COMMENTS
The authors have addressed all the questions and suggestions made in the second review. I found only minor corrections

needed in Fig. 1, along with a remark that the authors could easily address in one or two sentences. Beyond that, I think the manuscript could be accepted for publication.

SPECIFIC COMMENTS
Figure 1a-d. In the y-axis labels, replace “alpah” with “alpha”.
Figure 1 caption: Make sure the subplots are properly referenced; for example, subplots a and b are said to feature shaded areas representing individual species, but these appear in a and c.

Lines 167-170: Is this result a consequence of a negative relationship between species richness and functional diversity related to the fast-slow leaf economic spectrum? Fig. 3a,c suggest so. If this is correct, please specify and indicate whether this correlation is expected or known.

(Remarks on code availability)
As in the former reviews, I did not run the code, as I cannot provide expert advice on R programming language. However, the code seems sufficiently commented and includes human-readable .csv files, a README.txt file describing the other files content and role. The necessary packages are installed when R code is executed and the libraries referenced in the manuscript are included.

Reviewer #1 (Remarks to the Author):

Major comments
This study is exceptional in the spatial extent and high replication of the underlying data, covering a large area of China. The collection of plant trait data and soil chemistry over such a geographic extent, I believe specifically for this study, is a remarkable achievement. Finally, this study is notable for its integration of field data and remote sensed NDVI data, and its use of piecewise structural equation models (pSEMs) to understand drivers of diversity and stability. I found the conclusions, including the association of fast strategy plants with stability, interesting and well supported by the pSEM. These conclusions will be a useful contribution to the nascent field of spatial scaling of biodiversity-stability relationships. I reviewed the deposited R code and found it well commented and comprehensive.

> Response #1-1: we are grateful for these positive comments on this manuscript and for your suggestions on how to improve it further.

1. I do have some concerns with the way alpha stability was calculated, the use of data dredging and backwards selection approaches, and the many different analyses applied to the same variable set. I detail these concerns below ("Methodological comments").

> Response #1-2: we have carefully considered the reviewer's valuable methodological concerns regarding alpha stability calculation, variable selection approaches, and analytical redundancy. Specifically, we (1) checked the formulation of alpha stability and revised the definition of alpha stability in the text; (2) removed the model selection statistics, (3) replaced the bivariate analyses with partial regressions that control for the effects of other variables to maintain the consistency with the SEM results. The revised manuscript has benefited significantly from these methodological enhancements, and we believe it now presents a more robust and transparent analysis.

2. The writing was generally smooth, but I found a few inconsistencies in the arguments presented, detailed below ("Writing comments").

> Response #1-3: we sincerely appreciate the reviewer's careful reading and suggestions regarding argument consistency. We have carefully addressed each concern through comprehensive revisions to the manuscript, as detailed in our point-by-point responses below.

3. In some cases, the authors did not find support for hypothesized relationships between functional diversity and stability, or between beta richness and spatial asynchrony. Here the authors could have usefully considered the possibility that the hypotheses were sound but applied either to traits not measured or to larger distances between the subplots. No rationale was given for the spatial distances between the subplots (relatively small, the centers of 30 x 30 m subplots were 50m apart – so the gap between subplots was presumably 20m): was there any empirical data suggesting that metacommunity processes like plant dispersal were

important in the range 20-50m for example?

> Response #1-4: we fully acknowledge the reviewer's observation that certain unmeasured hard functional traits (e.g., root architecture, hydraulic conductivity, or secondary metabolite production) may indeed play critical roles in mediating diversity-stability relationships, potentially explaining some of the non-significant findings in our study. As highlighted by Carmona et al. (2021), belowground traits, particularly those related to fine roots, may have disproportionate effects compared the aboveground traits, as many species sharing similar fine-root syndromes exhibit contrasting aboveground traits. Yet, in large-scale studies, how belowground traits affect ecosystem stability due to measurement challenges remains underrepresented. We have added this limitation to the Discussion (L245-251).

Regarding the spatial scale of our subplots (30 × 30 m with 50-m spacing), our design was informed by empirical evidence from comparable grassland ecosystems. Synthesis studies showed that the seed dispersal distances of herb and graminoid can be from 0 to up to hundreds of meters (Bullock et al. 2017; Lososová et al. 2023), with a rapid decrease in dispersal likelihood with increasing distance (Bullock et al. 2017). The 50-m plot size aligns with the sample area requirements for capturing grassland community structure in the field, as validated by Wang et al. (2014) in similar ecosystems. We agree that although dispersal between our subplots may have occurred, the relatively short distance could have led to high similarity in species composition, potentially limiting the detection of strong beta diversity and spatial asynchrony patterns. We have also discussed these points in the Discussion (L252-259).

Methodological comments

4. Methods: "Alpha stability of subplots was defined as the weighted average of community stability across subplots." Please explain what you mean by weighted? Following your formulae, it looks like you summed the NDVI-based productivity value across all four subplots, then summed the standard deviation of NDVI productivity across all four subplots, and then took the ratio of these values. However, I think you should have instead calculated the mean/sd value (inverse cv) for each subplot separately and then averaged these four values. Alpha stability is the denominator of spatial asynchrony so this calculation will affect this metric too. I agree with the calculation of gamma stability.

> Response #1-5: we calculated gamma stability, alpha stability and spatial asynchrony following the method from Wang et al. (2019, pp5), but differently using 1/CV as temporal stability. In their method, they defined the variability of each local community k as

$CV_{C,K} = \sqrt{v_{\sum,kk}}/\mu_{\sum,k}$, where $\mu_{\sum,k} = \sum_i \mu_{i,k}$,

i.e., the temporal mean of total community biomass in patch k, calculated by summing the temporal mean biomass of all species i within the patch; and

$v_{\sum,kk} = \sum_{i,j} v_{ij,kk}$,

i.e., the temporal variance of total community biomass in patch k, considering the covariance $v_{ij,kk}$ among all species i, j in the patch.

Local-scale average community variability (or alpha invariability) was defined as the weighted average of community variability across patches,

$$CV_{c,l} = \sum_k CV_{C,K} \times \frac{\mu_{\sum,k}}{\mu_{\sum,\sum}}$$

where $\mu_{\sum,\sum} = \sum_k \mu_{\sum,k}$ is the temporal mean of total biomass across all patches; the weighted term $\frac{\mu_{\sum,k}}{\mu_{\sum,\sum}}$ ensures that patches with larger biomass contribute more to the overall variability.

By substituting the expression for $CV_{C,K}$, we have:

$$CV_{c,l} = \sum_k \frac{\sqrt{v_{\sum,kk}}}{\mu_{\sum,k}} \times \frac{\mu_{\sum,k}}{\mu_{\sum,\sum}} = \sum_k \frac{\sqrt{v_{\sum,kk}}}{\mu_{\sum,\sum}} = \frac{\sum_k \sqrt{v_{\sum,kk}}}{\mu_{\sum,\sum}}$$

where the numerator $\sum_k \sqrt{v_{\sum,kk}}$ represents the sum of the temporal standard deviation of total biomass across all patches (i.e., the total biomass variability across all patches). The denominator $\mu_{\sum,\sum}$ is the sum of the temporal mean total community biomass across all patches (i.e., the total mean biomass across all patches).

Overall, $CV_{c,l}$ reflects the relative variability of community biomass at the local scale, quantifying how much biomass fluctuates in relation to its mean. However, this cannot be recognized from the final formula. To be more specific, we revised this sentence as 'Alpha stability of subplots was defined as the average of community stability across subplots, weighted by the mean total biomass of each subplot' in the text.

5. Statistics. I am puzzled why the authors include individual linear model analyses (gls) in addition to the piecewise SEM. The pSEM is superior to all individual linear model approaches as it explicitly allows for a causal structure, whereas the linear models run the risk of including spurious variables or excluding indirect effects. Further, it literally is data dredging to use the dredge function in MuMin, where all possible models are run, relative importance of each variable is assessed and model selection is used to choose the most efficient model. Statisticians specifically counsel against data dredging approaches.

> Response #1-6: we agree that pSEM is a more powerful approach for testing hypothesized causal structure and identify both direct and indirect effects. However, individual linear models can provide a more detailed view of specific relationships. For example, previous Fig. 3a and Fig. 4c show a larger range of spatial asynchrony in QTP than IMP, which cannot be reflected by pSEM. The individual linear models can also verify the robustness of certain paths in pSEM, e.g., the positive relationship between alpha richness and alpha stability, and no relationship between beta richness and spatial asynchrony in both regions. Therefore, we kept the individual linear models as a complement for pSEM. However, we agree with the reviewer that the individual linear models run the risk of including spurious variables. To avoid this problem and obtain more consistent conclusions with the SEMs, instead of using bivariate analyses, we estimated the partial effects of alpha richness, CWM PC1 and alpha RaoQ on alpha stability, and the partial effects of beta richness and beta RaoQ on spatial asynchrony, respectively, while controlling all other variables included in the SEM model. We have revised the text accordingly.

To avoid data dredging, we removed the model section approach throughout the text. Instead, we followed the reviewer's suggestion (Response #1-7 below) and calculated the total effect of an explanatory variable on a response variable (including direct and indirect effects). This approach can also reflect the most influential factor(s) for alpha stability or spatial asynchrony.

6. The a priori SEM is a series of hypotheses about how the system works. I disagree with removing non-significant paths in pSEMs to improved the model fit (as judged by Fisher's C?) as this would be altering your a priori hypotheses. If a path is not significant in a pSEM, it may still be affecting the overall model, for example via indirect paths and hidden covariance. By imposing a backwards selection approach on their pSEMs, the authors forego their hypothesis-based approach. Backwards selection is another approach generally discouraged by statisticians. The total effect of an explanatory variable on a response variable, including indirect effects, can be calculated by the semEff package which uses bootstrapping to assess significance; this may be a better way to summarize your model.

> Response #1-7: we agree that removing non-significant paths in pSEMs to improve the model fit (as judged by Fisher's C) could risk altering the a priori hypotheses. To avoid this, we followed the reviewer's suggestion and kept all the non-significant paths in the model (Fig. 4; non-significant paths were not shown in the figures).

To calculate the direct, indirect and total effects of an explanatory variable on a response variable, we have tried the semEff package, but unfortunately our gls models with the correlation structure (correlation = corRatio(form = ~ LAT + LON)) cannot be applied to bootstrapping procedure in 'bootEff' function of the current version, and the result from 'multigroup' function cannot applied to 'bootEff' function either (We have verified this limitation with the author of this package, Mark V. Murphy). Therefore, we used custom R code to calculate direct, indirect effects and total effects of variables for the multigroup SEM

by multiplying the standardized coefficients of the involved paths and summing all resulting products. The R codes are available via Figshare at https://figshare.com/s/58d960e0f8ed868b0524. We also provided this information in the Method section.

Writing comments.

7. At first reading, these two sentences in Abstract seem contradictory: "In contrast, we found no evidence that variation in species richness and functional diversity among local communities contribute to stabilize productivity at larger spatial scales. Although the effects of environmental conditions on local diversity and stability differed between the two regions, the relationships between diversity and stability remained consistent at both the local and larger spatial scales." After reading the manuscript I understand the results that you are trying to convey, but the abstract should be understandable at first reading.

> Response #1-8: we followed this suggestion and revised these sentences to be clearer as "In contrast, we found no evidence that variation in species richness and functional diversity among local communities contributes to stabilize productivity across larger spatial scales. While the relationships of environmental conditions with local diversity and stability differed between the two regions, the overall positive relationships between diversity and stability were consistent at both the local and larger spatial scales." (L34-39).

8. Introduction. Line 54: "First, resource acquisition strategies determine species' responses to environmental changes, with fast species exhibiting faster recovery but lower resistance and slow species having higher resistance but slower recovery. In this context, while communities dominated by fast species can exhibit low community stability due to low resistance, communities dominated by slow species can exhibit high community stability due to high resistance ". This seems contradictory: First sentence establishes fast species have resilience and slow have resistance, yet second sentence says that only slow have stability as they are resistant (but resilience is also a type of stability). I think here the issue is that the literature perhaps emphasizes slow strategies and resistance, but in your Discussion you propose that fast strategies lead to resilience that also stabilizes productivity. The Introduction is perhaps caught between these two ideas, but you need to either set up two alternate hypotheses at the start (one resistance based, one resilience based) OR you need to restrict the Introduction to describing only the more common hypothesis.

> Response #1-9: thank you for this helpful comment. We agree that the original text created confusion by emphasizing resistance-based stability while mentioning resilience only briefly. We have revised the Introduction to clarify that communities dominated by fast species (due to high resilience) and slow species (due to high resistance) can both exhibit high community stability. We now present these as two alternate hypotheses (L54-57), which better aligns with the ideas discussed later in the Discussion.

9. Line 60: “species can either respond asynchronously to environmental fluctuations when they differ in their functional traits (Fig. 1a2), or display competition-induced negative covariance if they have similar traits (Fig. 1a3). Therefore, communities with low or high functional diversity in fast-slow traits are likely to show greater stability due to compensatory dynamics or negative covariance” This seems contradictory: Greater stability due to compensatory dynamics or negative covariance is, according to the first sentence, only possible under high diversity, not low diversity (contradicts second sentence).

> Response #1-10: thank you for pointing out this inconsistency. We agree that the original wording may have been confusing. While high functional diversity can promote stability through compensatory dynamics driven by asynchronous responses to environmental fluctuations, low functional diversity can also lead to stability through negative covariance arising from biotic interactions among functionally similar species (de Bello et al. 2021). To clarify this, we have revised the sentences as “negative covariance due to competition” (L60) to be clearer. This change helps distinguish the different mechanisms by which both high and low functional diversity can enhance community stability.

10. Figure 1 needs much more explanation in the legend, describing what the different colours represent, what all the abbreviations mean, what level the traits are calculated at (species? community?) in each part of (a), and how part (a) and (b) relate to each other. It is also not stated what the trait is, as there are many traits where a high CWM (community weighted mean) trait value would not be expected to lead to high alpha stability. The authors are making their argument explicitly for the fast-slow trait continuum, with CWM of the “slow” strategy argued in the Introduction to lead to resistance. Even having read the paper, I am struggling to understand the arguments presented in this figure, suggesting that it should also be better explained in the main text.

> Response #1-11: we followed this suggestion and added more explanation in Figure 1. Specifically, in Figure 1a-b, ‘Trait distributions filled in different colors represent the two different species in a community. The lines in red and blue represent two different communities; the bolder lines represent communities at time t_0, while the thinner lines represent the same communities at time t_1.’ In Figure 1c-d, ‘The red and the blue lines represent the trait distributions of two communities in a metacommunity that depicted in (a) and (b) respectively’.

The figures show the trait distributions related to fast strategy traits (including specific leaf area, leaf N and P content). Based on the updated hypotheses in the Introduction, we revised the figures and show that both fast and slow strategy CWM could lead to high alpha stability to be clearer.

11. Statements such as “a positive trend between CWM and alpha stability” are non-informative without knowing what traits loaded positively on the underlying PCA axis. I suggest modifying CWM throughout the manuscript to clarify the direction, e.g. “fast strategy

CWM".

> Response #1-12: we followed and modified 'CWM' without a specific direction throughout as 'fast strategy CWM'.

12. The logic behind this statement (Line 225) is unclear. Why would the effect of diversity on stability increase with mean diversity? "In our study, Qinghai-Tibet Plateau has locally higher species richness (mean = 12, se = 0.67; Fig. S7a) than Inner Mongolia Plateau (mean = 9, se = 0.36) or the regions analyzed in Ren et al.42 (mean = 8), making alpha richness a more important predictor of alpha stability than the environment."

> Response #1-13: thank you for this insightful comment. We agree that our statement needs more explanation. We agree that the absolute level of species richness (mean = 11 vs 7) doesn't have a direct causal link to a stronger diversity-stability relationship. However, this pattern can arise because higher richness increases the probability of compensatory dynamics and functional redundancy, which can enhance the stabilizing effects of biodiversity. We revised this sentence in the text in L229-232 as '…Alpha richness a more important predictor of alpha stability than the environment in Qinghai-Tibet Plateau probably because that greater species richness allows for stronger compensatory dynamics and functional redundancy that stabilize community biomass'.

Minor comments

Line 96 change "along" to "among".

> Response #1-14: Done.

Reviewer #1 (Remarks on code availability):

Yes I reviewed the R code. The code was linearly arranged and well commented. The link between code and figures was particularly well done. The readme provided a description of all variables in the data set (i.e. meta-data) and information on executing the code.

> Response #1-15: thank you for your positive feedback on the code structure and documentation. For this new version, we have updated some parts of the code to reflect the changes suggested during the review process. Please, find it here: figshare.com/s/58d960e0f8ed868b0524.

Reviewer #2 (Remarks to the Author):

SUMMARY

This study examines the relationships among species and functional diversity, and the stability of biomass productivity within plant communities at local (alpha-level) and larger (beta-level) spatial scales. Using data collected in 237 grassland sites in the Qinghai-Tibet Plateau (93 sites) and the Inner Mongolia Plateau (151 sites), the authors assess more specifically the relationships between alpha- and beta-species richness, alpha- and beta-functional composition and diversity, and alpha-biomass productivity stability and spatial asynchrony in biomass productivity. Using Structural Equation Modelling (SEM), the authors further tested the direct and indirect effects of environmental factors (i.e., temperature, precipitation, and soil fertility) on the stability and spatial asynchrony of biomass productivity through their impacts on species richness and the functional composition of plant communities. The results show that species richness and the presence of fast-growing species positively influence the stability of biomass productivity at the local level. In contrast, no relationship was found between beta-richness, beta-functional diversity, and the spatial synchrony of biomass productivity. These findings hold true in both regions; however, the authors further demonstrate that the pathways through which environmental factors influence alpha-biomass productivity and spatial asynchrony in biomass productivity differ between the two regions.

While this study relies on an impressive dataset, I have major concerns regarding this manuscript:

> Response #2-1: We sincerely appreciate the reviewer's thoughtful engagement with our study. We fully acknowledge these important concerns and will address each point systematically in our revision.

1. Clarity of terminology: The manuscript is very hard to follow due to imprecise use of terminology. For example, the authors use the term "community stability" throughout the Introduction. To me, "community stability" typically refers to the stability of species richness or species abundance (see, e.g., Jarzyna et al. 2022; Ghosh et al. 2024). However, I only realized at the end of the Introduction that "community stability" in this study actually refers to the stability of ecosystem functioning, as captured by plant biomass productivity. Similarly, in Fig. 1, the authors hypothesize that communities with high community-weighted mean (CWM) and low alpha functional diversity, with low CWM but high alpha functional diversity, or with both low CWM and low alpha functional diversity can lead to high alpha stability. However, it is unclear which trait is considered here.

> Response #2-2: we thank the reviewer's insightful comments regarding terminology clarity, which have revealed important opportunities to strengthen our manuscript's conceptual rigor. We have clarified that we focused on 'temporal stability of community productivity' in the Abstract, and also defined 'community stability' as 'the temporal stability of community

biomass productivity' when it first appeared in the Introduction (L47). We consistently refer to 'community stability' when discussing temporal stability at the local community level. Because in our study we also considered the temporal stability at the metacommunity level (which includes four communities), we used the term 'ecosystem stability' (mainly in the Discussion) when examining stability patterns across different scales. Where appropriate, we also used 'temporal stability of productivity' directly to avoid ambiguity. The precedents can be found in previous papers related to stability across spatial scales (Wang et al. 2019; Hautier et al. 2020).

We have added the details of the functional traits used (i.e., specific leaf area, leaf nitrogen and leaf phosphorus content) in the legend of Figure 1, and also in the Introduction (L97-98) to be clearer.

For instance, the results would likely differ depending on whether the trait is SLA or LDMC. Other examples of ambiguous terminology are provided in the specific suggestions below. Overall, I believe the manuscript requires clearer definitions and justifications for the metrics used. These metrics should be explicitly defined in the last paragraph of the Introduction, with references to the Methods section for further details. Additionally, I suggest clearly defining the metrics and abbreviations in Figure 1 to enhance readability and understanding.

> Response #2-3: we appreciate the reviewer's valuable suggestions, which greatly improved the clarity of our terminology and trait-specific analyses. We followed this suggestion and have added more details to specify that the traits are related to fast strategy including SLA, leaf N and P content (L97-98). The revised Figure 1 now includes clearer annotations with trait metrics and acronym definitions.

2. Justification of Structural Equation Modelling (SEM): The authors used Structural Equation Modelling and provided a conceptual figure (Fig. 1) to illustrate the relationships to be tested. While these analyses are well-conducted, I believe the structure of the model should be better justified. I found the manuscript lacking sufficient explanation regarding the links included in the conceptual framework. This information could perhaps be added to the Methods section, included in the legend of Figure 1, or summarized in a table that could be placed in the Supplementary Materials. From my point of view, it is crucial to demonstrate that the SEM structure is based on solid knowledge and strong hypotheses (see Grace 2008).

> Response #2-4: we sincerely appreciate the reviewer's suggestion regarding the need to strengthen the theoretical foundation of our SEM framework. In response, we have significantly expanded the methodological justification in several key ways:

First, we have added the hypotheses for each path based on available theoretical and empirical findings from previous research in Supplementary Methods 2. This includes clear a priori hypotheses for the relationships between environmental drivers, diversity components, and stability metrics.

Second, we have added more details about the ecological rationale in the Method (L425-428) for the main paths in our SEM model.

These modifications collectively address the crucial need to demonstrate that our SEM structure rests on solid theoretical foundations while maintaining the manuscript's flow.

3. Scale mismatch in data collection: I also have concerns about the mismatch between the scale at which data on plant diversity were collected (i.e., in 4 subplots of 0.5 m × 0.5 m) and the scale at which data on biomass productivity were collected (i.e., 4 subplots of 30 m × 30 m). Although these subplots overlap, such a scale mismatch in data collection might explain why beta-diversity and beta-functional diversity are not influencing the spatial asynchrony of biomass productivity. Since the plant diversity data were collected at a relatively small spatial scale and not upscaled to larger scales, I also suggest toning down the strong conclusions regarding diversity-stability relationships at large spatial scales.

> Response #2-5: we appreciate the reviewer's comment about the scale mismatch between plant diversity and biomass productivity data, and we fully acknowledge this limitation. Due to the significant time and labor demands of large-scale plant diversity monitoring, and in line with many other published studies, we relied on biomass data derived from remote sensing, acknowledging its inherent resolution limitations (García-Palacios et al. 2018: 30 m *30 m in the filed vs 250 m * 250 m pixel size; Oehri et al. 2017: 1 km^2 filed plot vs. 250 m * 250 m pixel size; Chen et al. 2021: 100 m * 100 m field plot vs. 250 m * 250 m pixel size; Wu et al. 2023: 100 m * 100 m field plot vs. 250 m * 250 m pixel size). Despite these apparent mismatches, previous studies have demonstrated that scale differences between field and remote sensing data do not necessarily pose a problem. For example, García-Palacios et al. (2018) showed that mean NDVI using a pixel size of 30 m × 30 m and the mean NDVI using a pixel size of 250 m × 250 m was strongly correlated (R^2 = 0.89), suggesting that remote sensing data at both resolutions can reliably capture spatial variation observed in the field. In response to this concern, we have explicitly discussed the potential effect of scale mismatches explaining the neutral relationship between beta diversity and spatial asynchrony in the Discussion section in L259-265.

4. Clarity of statistical analyses: Finally, it was not always clear to me, and it was not well explained in the Methods section, why the different statistical analyses were conducted. For example, I do not understand the rationale for using both an ANCOVA and a model selection analysis to test the effects of the different variables on alpha stability and spatial asynchrony. How do these analyses complement each other, and do they not yield very similar results? Providing a clearer justification for the use of these methods and explaining their distinct contributions to the analysis would significantly enhance the clarity and robustness of the manuscript.

> Response #2-6: we thank the reviewer for raising this important methodological

consideration. In response to this comment and also the similar comments from Reviewer 1 (Response #1-6), we have streamlined our analytical approach by removing the model selection analysis. We retained structural equation modeling as our primary statistical framework in the revised manuscript, as this more parsimonious approach better aligns with our study objectives while maintaining analytical rigor. We also replaced the ANCOVA with partial regressions after controlling the effect of other variables. Partial regressions not only visually display the distribution of data points, but are more consist with the results from SEM. We have updated the Methods and the Results sections accordingly. We also provided clearer justification that the contribution of structural equation model was to assess the complex relationships between stability components and their drivers in our study system (L418-423). And partial regressions were used to estimate the partial effect of diversity on stability, while controlling for the effects of all other variables included in the model (L436-439).

Below, I present more specific suggestions that may help, I hope, to improve the manuscript.

SPECIFIC SUGGESTIONS

•L39-41: This section is unclear. What is meant by "environmental change"? Please clarify or provide examples to help the reader understand.

> Response #2-7: we thank the reviewer for this comment and have revised "environmental change" to "environmental factors (e.g., precipitation)" to be clearer.

•L51: Does "community stability" here refer to the stability of species richness over time, or do you mean the stability of biomass production? To me, "community stability" typically refers to the stability of abundance or diversity within a community (see e.g. Jarzyna et al. 2022; Ghosh et al. 2024), rather than the stability of ecosystem functioning or biomass production.

> Response #2-8: we thank the reviewer for raising this important clarification. We confirm that in our study, "community stability" specifically refers to the temporal stability of plant community biomass productivity, not species richness or composition stability. As our Response #2-2, we have explicitly defined this terminology in both the Abstract (L27-28) and Introduction sections (L48-49) and have maintained consistent usage throughout the text to avoid ambiguity. We also cited related literature that calculate the temporal stability of biomass productivity (Wang et al. 2019; Hautier et al. 2020) in the Method section.

•L54-59: It is hard to see how the mechanisms discussed are related to or reflected in the conceptual framework supporting the SEM presented in Fig. 1. Could you provide a clearer explanation or better link these ideas?

> Response #2-9: we have added the hypotheses in Supplementary Methods 2 for each path in the conceptual SEM in Fig. 1, based on available theoretical and empirical findings from previous research.

•L72-75: Apologies for pointing this out, but this section is also unclear. Does "temporal stability" here refer to the stability of biomass production, species richness, or something else? Please specify.

> Response #2-10: it refers to the temporal stability of plant productivity. We have revised 'temporal stability' to 'temporal stability of productivity' throughout to be clearer.

•L81: I suggest replacing "essential" with "a strong driver of" for greater precision.

> Response #2-11: replaced.

•L98: Once again, the term "community stability" is ambiguous. Please clarify its meaning in this context.

> Response #2-12: here we revised "community stability" to "temporal stability of productivity" here, also when examining stability patterns across different spatial scales, to be more rigorous.

•L101-103: Why do you expect the pathways through which environmental factors affect stability to differ between the two regions? Providing a stronger rationale or referencing relevant studies would help strengthen this point.

> Response #2-13: The expected differences in environmental pathways between regions stem from their distinct limiting factors. Precipitation is recognized as the most important limiting resource at the Inner Mongolia Plateau, while temperature is the most important limiting resource at the Tibetan Plateau with high elevation, as evidenced by regional ecological studies (e.g., Ma et al. 2010). This difference can fundamentally alter how environmental factors mediate stability mechanisms as supported by other studies related to stability (e.g., Bai et al. 2004; Wang et al. 2020; Ma et al. 2017). We have strengthened this rationale in the revised manuscript by incorporating relevant references and clearer ecological explanations for these contrasting dynamics (L91-94).

•L107: It only becomes clear here that your focus is on the stability of productivity. Consider clarifying this earlier in the manuscript to improve the reader's understanding.

> Response #2-14: we have followed this suggestion and clarified it as 'stability of productivity' in both Abstract and the beginning of the Introduction.

•L121-130: Why are you using both an ANCOVA and a model selection analysis? How do these analyses complement each other? Clarifying their distinct purposes and contributions would be helpful.

> Response #2-15: In response to this suggestion and also the suggestions from Reviewer 1 (Response #1-6), we have streamlined our analytical approach by removing the model selection analysis while retaining structural equation modeling as our primary statistical framework. We also replaced the ANCOVA with partial regressions after controlling for the effects of other variables, as explained above (Response #2-6).

•L140-143: This sentence is unclear. In the SEM presented in Fig. 1, there is no direct link between spatial asynchrony and alpha stability, so beta diversity cannot affect alpha stability via spatial asynchrony. Is that correct? Please revise for clarity.

> Response #2-16: we appreciate the reviewer's careful observation. We have added the hypotheses in Supplementary Methods 2 for each path in the conceptual SEM in Fig. 1e, and showed that there is no causal link between spatial asynchrony and alpha stability. We have revised this sentence to be clear as "…alpha diversity (both alpha richness and functional diversity) consistently contributed to alpha stability in both regions, while beta diversity (both beta richness and functional diversity) did not contribute to spatial asynchrony…" in L123-126.

•L156-157: Is it shown somewhere that higher mean annual precipitation and soil fertility contributed to higher spatial asynchrony and higher gamma stability? If so, please provide a reference or elaborate.

> Response #2-17: in our updated results, only higher mean annual precipitation contributed to higher spatial asynchrony. We have revised this sentence as "…higher mean annual precipitation directly promoted spatial asynchrony without influencing beta richness or beta RaoQ" to be clearer. Previous studies hypothesized in their *a prior* SEM that mean annual precipitation can affect spatial asynchrony directly or indirectly by affecting plant diversity (Qiao et al. 2022; Song et al. 2024). However, their results didn't find a significant relationship between mean annual precipitation and spatial asynchrony. We also have added these hypotheses in Supplementary Methods 2 for these paths to support our results.

•L171-172: You did not assess ecosystem stability at large spatial scales; your analysis was limited to the alpha level for stability if I understood correctly. I think that this conclusion is not supported by your results and appears to be beyond the scope of the study.

> Response #2-18: we appreciate the reviewer's comment. We would like to clarify that, in our study, temporal stability is assessed at both the local scale (alpha stability) and at larger spatial scales (gamma stability). We have applied the method developed by Wang et al. (2019) to explicitly partition the gamma stability at larger spatial scales (the stability of the metacommunity) into alpha stability (local community stability) and spatial asynchrony (asynchronous dynamics of productivity across local communities). However, we didn't include gamma stability in our analyses mainly because its strong correlation with alpha stability and spatial asynchrony. But beta richness, beta functional diversity and spatial asynchrony are measuring diversity and asynchronous dynamics of productivity among

communities at larger spatial scales.

Based on our updated analyses, we have revised our conclusion and also this sentence as “our study reveals that alpha diversity, rather than beta diversity, influences ecosystem stability at larger spatial scales”.

•L187: This finding is not related to CWM but rather to the ‘slow-fast’ PCA axis.

> Response #2-19: we have revised it as “CWM PC1 representing fast strategy traits”.

•L204-206: How was this classification performed? I could not find this information in the Methods section. Please provide details.

> Response #2-20: we used the data from the China Terrestrial Ecosystem Database (https://www.ecosystem.csdb.cn/index.jsp), which is described using six hierarchical levels, based on the plant community classification system proposed in Vegetation of China (ECVC, 1980) and referring to the classification approach in China Ecosystem (Sun et al. 2005). We have now added the detailed classification methodology and reference in the Supplementary Methods 3 as a support.

•L211-212: This contradicts previous studies, such as García-Palacios et al. (2018). Consider discussing how your findings align with or diverge from these studies.

> Response #2-21: thanks for the reviewer’s suggestion. While García-Palacios et al. (2018) found a positive relationship between species richness and ecosystem stability in high aridity sites at the global dryland scale, we observed a negative relationship between alpha diversity (both richness and functional diversity) and alpha stability specifically in the desert steppe vegetation type. We believe this difference may reflect a context dependency in the biodiversity–stability relationship. García-Palacios et al. (2018) analyzed broad-scale, stratified by arid conditions across multiple dryland vegetation types and continents, whereas our analysis focused on smaller scale, stratified by vegetation type. This finer resolution allows us to detect habitat-specific mechanisms that might be masked at broader scales.

On the other hand, García-Palacios et al. (2018) showed a positive relationship between species richness and ecosystem stability in high aridity sites, they also showed that as aridity increased, the link between the variance of SLA and stability shifted from positive to negative. Their results suggest that strong environmental filtering at high aridity may limit the stabilizing effects of functional diversity and shift the dominant mechanism to functional redundancy. This is essentially consistent with our interpretation, i.e., high alpha diversity may reflect patchiness rather than effective species interactions, potentially explaining the negative relationship with stability. We have extended the text and discussed the potential differences between García-Palacios et al. (2018) and our study.

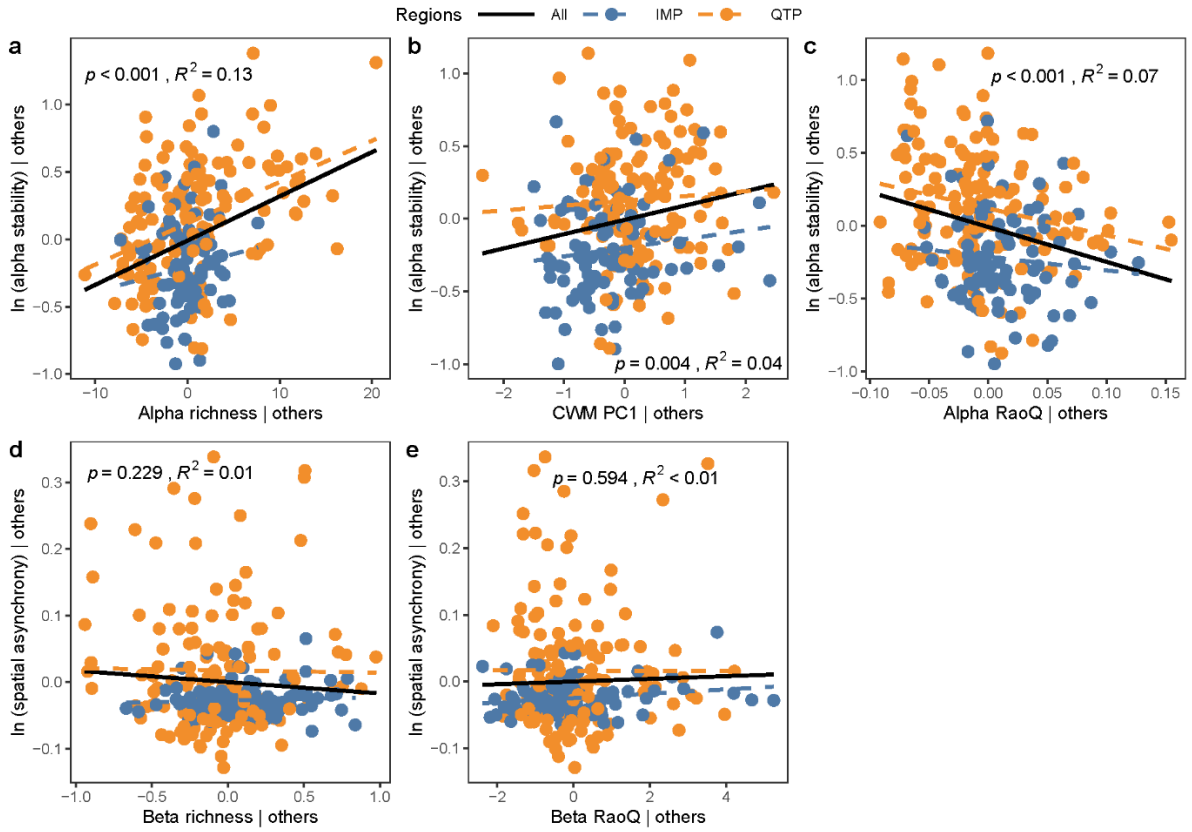
•L268-269: This conclusion is only valid at the local scale, correct? If so, please specify to avoid overgeneralization.

> Response #2-22: we have revised our conclusion based on the updated results as “In conclusion, our results demonstrate that while species richness and functional diversity contribute to alpha stability, they show no relationship with spatial asynchrony”.

•L285: Did you account for the sampling year in your statistical analyses? If so, please specify how.

> Response #2-23: we agree that the sampling year could potentially affect functional traits, species richness and soil fertility sampled in the field. However, each site was only sampled once, which prevents us from directly testing whether these variables differed between the two years.

To test if sampling year affected the relationships of interest, we re-ran the same analyses described in the manuscript, adding 'year' as a predictor in the GLS models in the SEM. The partial regression plots (below) show no substantial differences compared to the original results presented in Fig. 3, indicating that including 'year' does not significantly alter the results. Moreover, as 'year' is not a primary focus of our study, we have retained the original results in the manuscript.



•L294: How did you record aboveground biomass? Adding more details would enhance the Methods section.

> Response #2-24: we recorded aboveground biomass by harvesting all vegetation at ground level within each of the four systematically placed 0.5 m × 0.5 m quadrats per site, with samples oven-dried at 65 °C to constant weight. We have now expanded the Methods section to include these standardized protocols in L307-310.

•L295: Replace "collected" with "measured" for precision.

> Response #2-25: replaced.

•L295-296: Did you analyze the Qinghai-Tibet Plateau and Inner Mongolia Plateau sites together? Are the species pools similar in these two regions?

> Response #2-26: for functional traits, we first calculated the proportion of species biomass measured for specific leaf area (SLA) to total aboveground biomass at each site, and then took the average value (89.8%) across all sites at the Qinghai-Tibet Plateau and Inner Mongolia Plateau. We added more details in this sentence.

For all the analyses in this study, we analyzed the Qinghai-Tibet Plateau and Inner Mongolia Plateau sites together and included a 'Regions' column with 'QTP' (Qinghai-Tibet Plateau) and 'IMP' (Inner Mongolia Plateau) as two levels in the dataset. More details are available in our data and R code at https://figshare.com/s/58d960e0f8ed868b0524.

There is a 10.2% overlap in species composition between the two regions. We have added this information in L311.

•L316: Specify what "temporal stability" refers to in this context.

> Response #2-27: the temporal stability of productivity. We have made this to be clearer in the revised manuscript.

•L341-342: Did you calculate the mean and variability of annual temperature and precipitation per year or per growing season?

> Response #2-28: we calculated the mean of temperature and precipitation using full annual climate data rather than growing-season subsets, as this approach better captures the complete environmental constraints on these perennial-dominated ecosystems. We have revised this sentence to be clearer.

•L351: Ref. 12 does not seem appropriate here. Did you mean Ref. 56?

> Response #2-29: thanks for pointing this out. Indeed, we have replaced reference 12 with Whittaker (1972).

•L360-361: This is unclear—weighted by what? Please clarify to avoid ambiguity.

> Response #2-30: we have revised this sentence as "Alpha stability of subplots was defined as the weighted average of community stability across subplots, weighted by the mean total biomass of each subplot". Please, see our response to Review 1 (Response #1-5) for more details.

•L394: Why did you use an ANCOVA? What was it used to test? How does it complement the model selection analysis? Providing a clearer rationale would improve the reader's understanding of your analytical approach.

> Response #2-31: in response to this comment and also the suggestion from Reviewer 1 (Response #1-6), we have streamlined our analytical approach by removing the model selection analysis while retaining structural equation modeling as our primary statistical framework. We also replaced the ANCOVA with partial regressions after controlling the effect of other variables, as explained above.

•L406: It would be interesting to examine the response of alpha mean productivity, not just stability. Including this analysis could provide additional insights to your study.

> Response #2-32: we appreciate the reviewer's thoughtful suggestion. In this study, we aim to assess how the environmental factors (temperature, precipitation, soil fertility) affect alpha stability and spatial asynchrony of plant aboveground productivity (the two mechanisms of gamma stability) mediated by plant diversity. We have recognized that some studies related to stability also tested the response of mean and standard deviation (SD) of the productivity beyond stability. However, while temporal stability is calculated as mean/SD, we think that analyzing the effects of drivers on mean or SD alone only provides a mathematical decomposition of stability. And examining the response of alpha stability and spatial asynchrony offers deeper ecological insight into the mechanisms underlying ecosystem stability responses to environmental change.

We agree that examining the response of mean productivity would be interesting. But this is beyond the scope of the current manuscript. We see it as a valuable direction for the future work.

•L419-431: Why was gamma stability not included in the SEM as a function of alpha stability and spatial asynchrony? This omission is puzzling, given its relevance in studies such as Wilcox et al. (2017), Hautier et al. (2020), Qiao et al. (2022).

> Response #2-33: we appreciate the reviewer's thoughtful suggestion. Indeed, gamma

stability as a function of alpha stability and spatial asynchrony in the SEM has been emphasized in several related studies, and we fully recognize its ecological and theoretical relevance.

We initially attempted to include gamma stability in the SEM as a function of alpha stability and spatial asynchrony. However, this specification led to a singular fit error when using gls() models with spatial correlation structures (corRatio(form = ~ LAT + LON)), which are essential for accounting for spatial autocorrelation in our dataset. This issue arises because gamma stability is mathematically dependent on alpha stability and spatial asynchrony (i.e., gamma stability = alpha stability × spatial asynchrony). After log-transformation, this becomes an exact additive relationship, resulting in multicollinearity that causes the model estimation to fail, particularly under the more stringent assumptions of gls.

More importantly, the central aim of our study is to understand how environmental drivers (i.e., climate and soil fertility) directly and indirectly influence the two components of gamma stability—alpha stability and spatial asynchrony—rather than partitioning the relative contribution of each component to gamma stability. Since we already know from prior theoretical and empirical work that gamma stability is shaped by both alpha stability and spatial asynchrony, our primary interest lies in how environmental drivers propagate through these components to ultimately affect gamma stability. Therefore, our modeling framework focuses on disentangling the pathways by which external drivers influence the underlying mechanisms of stability, which aligns with our research questions.

•Fig. 1: This figure is difficult to read, and some key information is missing. For example, please define terms like PC1, t0, t1, comm1, and comm2 in the caption. Additionally, the phrase "high community-weighted mean" is ambiguous. A high CWM SLA is not equivalent to a high CWM LDMC, and these measures cannot be interpreted in the same way. Specify which CWM is being referred to here.

> Response #2-34: we thank for these constructive suggestions. We have completely redesigned the figure and added more information in the figure legend to improved clarity. Specifically, (1) we have explicitly defined all abbreviations (PC1 = the first principal component axis; t_0/t_1 = time points, comm1/2 = two different communities). (2) We have specified that the CWM here is related to the fast strategy traits including specific leaf area, leaf nitrogen and leaf phosphorus content used in our study.

•Fig. S9: The labels on the x- and y-axes are not readable. Please adjust the font size or formatting to ensure they are clear and legible.

> Response #2-35: thanks for this suggestion. To response this comment and also the related comments from Reviewer 3, we have revised Fig. S5 (previous Fig. S9) with a density histogram to be clearer. Please also see the response to Reviewer 3 below (Response #3-12).

Reviewer #2 (Remarks on code availability):

The analyses and results of the paper are reproducible and the code can be easily used by the community. A 'README' file is provided and it contains instructions in order to reproduce some of the analyses, figures and results. I however noticed in the code that in some analyses, the variables were log-transformed. This could be detailed in the Methods' section. Note also that all the code is not available, for example I couldn't find the code used to run the model selection analysis and the SEM.

> Response #2-36: thanks for your suggestions. To improve the normality, alpha stability and spatial asynchrony were logarithm transformed. We have added this detail in the Methods section in L452-453. In the updated analyses, we have removed the model selection analysis, and also reorganized the structure and annotations of the code to make each analysis easier to locate and follow. The updated the codes is available at https://figshare.com/s/58d960e0f8ed868b0524.

Reviewer #5 (Remarks to the Author):

GENERAL COMMENTS

This manuscript assesses the role of plant taxonomic and functional diversity estimated from field sampling on the stability of productivity (estimated from remote sensing data) in multiple plots located in two different regions of different climates and biodiversity. The analyses take place at two spatial scales (local and regional), comparing large grassland-populated áreas with different environmental conditions.

The work and results fit the scope and relevance of Nature Communications. They are relevant for the ecological community and can be interesting for the remote sensing communities, where there is a growing interest in biodiversity and its relationship with ecosystem functions. Understanding the role of different facets of plant diversity on the stability of ecosystem functions is relevant for assessing the impacts of biodiversity loss, decision-making, and resilience management of natural and anthropic ecosystems. While not the first ones, the work uses a combination of ecological surveys and remote sensing data; the second assesses ecosystem function stability, which requires a long time series of continuous data unavailable from field surveys.

> Response #3-1: we appreciate the reviewer's positive assessment of our manuscript. We fully agree that combining field-based biodiversity data with remotely sensed productivity is a powerful approach to understanding ecosystem stability, especially across contrasting regions and spatial scales.

While the presented results support the conclusions, I have some concerns regarding the suitability of the remote sensing metric used as a proxy of ecosystem productivity. This choice might explain some of the results that disagree with those of former studies. New enhanced and more robust remote sensing proxies of ecosystem productivity might cast more reliable results. In particular, the use of remote sensing NDVI as a proxy may be outdated, particularly when looking at short-term temporal scales. Recent advances in remote sensing, such as the NIRvP index, would offer more accurate proxies of productivity and its stability. NIRvP combines an enhanced indicator of vegetation state (particularly in disturbed and semi-arid sites) with information about the incoming radiation, which might differ across sites. Some of the results presented regarding spatial asynchrony that contradict previous findings might result, in fact, from the use of NDVI, which is sensitive to bare soil and unable to account for across-site differences in incoming radiation.

> Response #3-2: thank you for recommending the NIRvP index. We followed this suggestion and calculated the NIRvP index. Please, see the response below for details (Response #3-16).

Also, maybe an important question to be at least clarified is “management”. While it is mentioned as a relevant element in the abstract (once) and the introduction sections (once), the

authors do not seem to consider the management regime in the sampling sites of this study. Since management can affect stability, are the sampling plots under different management regimes, and if so, are these known or unknown? If unknown, could these have had any significant impact on the results of this study? If known, how have these been accounted for?

> Response #3-3: we appreciate this thoughtful question. We agree that land management can influence ecosystem stability and appreciate the opportunity to clarify. In our manuscript, "management decisions" in the abstract and introduction refers to broader-scale land use and policy-level considerations that can be informed by biodiversity–stability relationships.

We acknowledge that you are referring to local management regimes at the plot level (e.g., grazing), which were not the primary focus of our analyses. However, we took steps during field site selection to minimize the influence of active land management. Specifically, all sampling sites were located away from settlements, roads, and regularly managed areas to reduce the impact of human disturbance and ongoing grazing. We carefully selected survey sites where vegetation showed no signs of livestock (cattle or sheep) browsing, indicating the absence of grazing until we did the survey. Field surveys were conducted annually between July and August, which corresponds to the peak growing season and a period when grazing is prohibited under local regulations. Specifically, according to policies such as the *Grass-Livestock Balance and Grazing Ban Regulations* of the Inner Mongolia Autonomous Region, grazing on natural grasslands is strictly banned from April 1 to June 30 each year. Furthermore, we avoided sites with substantial litter accumulation, as this would indicate a lack of winter grazing or mowing activities. We also intentionally selected sites that best represented the local natural vegetation to capture biodiversity and productivity under minimal direct management influence. We have added more details in the Method section in L301-303 to be clearer.

The description of the methods could also be improved, and some questions that could potentially affect the results were clarified.

> Response #3-4: we thank the reviewer for this constructive suggestion. We have thoroughly revised the Methods section to provide clearer experimental protocols and analytical procedures.

Finally, notice that Nature Communications is multidisciplinary; some concepts should be more clearly presented to the readers, which could not be deeply specialized in the topics of the article, despite their interest.

> Response #3-5: thank you for pointing this out. We have revised the manuscript thoroughly to improve the clarity of key concepts and ensured that the terminology and explanations are accessible to a multidisciplinary audience. For example, we have clarified the terms used in Fig. 1 to enhance understanding.

SPECIFIC COMMENTS

Introduction
Line 62: maybe clarify what you mean by “competition-induced negative covariance”.

> Response #3-6: we have revised it as ‘negative covariance due to competition’ to be clearer.

Lines 63-65: As phrased, this statement is confusing. Is it trying to summarize both mechanisms or only the second? Please summarize more clearly which conditions lead to each situation (stability or instability)

> Response #3-7: thanks for the reviewer’s comments. We tried to summarize the two conditions based on the compensatory dynamics in the second mechanism. We have revised these sentences as “…communities with high functional diversity in fast-slow traits are likely to show greater stability due to compensatory dynamics, while communities with low functional diversity may also be stabilized due to negative covariance” in L61-63 to be clearer.

Lines 72-75: Since alpha and gamma stability are defined as a sort of coefficient of variation of NDVI, how can then their

> Response #3-8: we sincerely appreciate the reviewer's important question regarding NDVI-based stability metrics and would be pleased to provide a more detailed response once the full comment is available. Based on the visible portion, we could clarify that: (1) we followed the framework proposed by Wang et al. (2019) to partition the gamma stability (temporal stability among communities at large spatial scales) into alpha stability (local community stability) and spatial asynchrony (asynchronous dynamics of productivity across local communities). (2) We assessed the correlation between field-measured biomass data from subset plots and NDVI from the remote sensing to make sure that NDVI can be a proxy as the plant aboveground productivity.

Line 99: Further specify what “larger spatial scales” are.

> Response #3-9: we specified both local and larger spatial scales as “from local (individual communities) to larger (an assemblage of four spatially separated communities) spatial scales”.

Figure 1: Fig 1 is confusing and should be improved.
1) First, the figure caption does not define all the terms included in the figure. What do “t1” and “t2” and the histograms associated represent? Different communities? Different species of the same community? Different traits? Different times? Please clarify.
2) Fig 1a and b represent the trait distributions but do not show the corresponding productivity stability or asynchrony; this would help the least specialized reader to understand what the authors mean by those terms. Also, the slow-fast spectrum is not represented or made explicit

in the plots. Should the reader understand that this spectrum is presented in the trait value axis? 3) Fig a1,3 do not seem to represent different CWM. As "(a1) high community weighed mean (CWM) and low alpha functional divergence"; however, the mean of the trait value seems to match the one of Fig. a3, defined as "both low CWM and low alpha RaoQ". Shouldn't the distributions of Fig. a1 be shifted towards the right side of the trait value axis? Fig. a1 only seems to present higher frequencies than Fig. a3, which, once normalized, do not modify the CWM.

> Response #3-10: we thank the reviewer for these constructive suggestions. We have redesigned Fig. 1 and revised the figure legend accordingly. Here are more details.

1) We have defined all the terms included in the figure in the figure legend. bold lines "t_0" represent communities at time t_0 in bold lines, while "t_1" represent the same communities at time t_1 in thin lines. Trait distributions filled in different colors in Fig. 1a-b represent two different species in a community. The lines in red and blue in Fig. 1a-d represent two different communities.
2) The y-axis of trait distributions represents the frequency, which can be considered as the weighted abundance of a species in the community. A higher peak height represents more dominant species. Combined with the trait value of each species, we can determine whether this community is dominated by fast- or slow-growing species, thus larger or lower community stability based on our hypotheses. The distance between two species represents the similarity in functional traits of these two species (functional diversity). We can also determine the functional diversity within and between communities, thus possible scenarios of spatial asynchrony based on our hypotheses. We have added this information in the figure legend. We have also specified that the CWM is related to fast strategy traits in the figure legend.
3) We have redesigned Fig. 1 to be more accurate. We used frequency as the weighted abundance of a species in the community. Now we have included four combinations of CWM and RaoQ as follows: (a) Community 1 with high CWM of fast strategy traits and low alpha RaoQ, and Community 2 with low CWM of fast strategy traits and low alpha RaoQ; (b) Community 1 with low CWM of fast strategy traits and high alpha RaoQ, and Community 2 with both high CWM of fast strategy traits and high alpha RaoQ.

Methods

Lines 285-286: The sample size of both regions is quite imbalanced; one is close to double the other. Could this induce any bias in the estimation of the regional diversities or variances? Have the authors checked that this does not modify the results, for example, by random resampling of equal sample sizes?

> Response #3-11: we followed this suggestion and conducted a sensitivity analysis in the multi-group SEM. We randomly selected 999 Qinghai-Tibet Plateau subsets matching the sample size of Inner Mongolia Plateau and refitting the model for each subset. The original coefficients fall within the 95% confidence interval of the 999 resampled estimates, so we consider the results to be robust to differences in sample size. We have added this analysis in the Methods section in L443-449 and also in Fig. S12-13.

Line 295: I'm surprised that sampling "five healthy, mature leaves 296 from up to five randomly-chosen individuals per species and site" covered almost 90 % of the sampled biomass. For plots with a few species and/or plants with numerous small leaves, five individuals might represent a small fraction of the sample's weight. How many species were there per plot? Fig. S9 is hard to read, and the presentation of absolute quantities makes it hard to understand whether, for example, plots with low taxonomic richness were systematically undersampled to measure their functional traits. I recommend making Fig. S9 clearer, which also includes absolute values of species richness and sampled biomass, and that the authors consider whether there were systematic biases in the characterization of plant traits.

> Response #3-12: we apologize for the confusion caused by our original description. To clarify, the figure of approximately 90% refers to the proportion of total site-level biomass accounted for by the species for which we measured functional traits, not the biomass of the sampled leaves or individuals themselves. In other words, the species included in our trait measurements collectively contributed about 90% of the aboveground biomass at each site. This ensures that our trait data represent the dominant biomass-contributing species. We have revised the Methods section to make this point clearer.

We appreciate the reviewer's comments regarding the number of species per plot and the clarity of Fig. S5 (previous Fig. S9). In our study, the community-level alpha richness for each site was the average species richness across the four subplots. And we used the total number of the species found in the four subplots as the gamma richness of each site. The species richness per site ranged from 1 to 37.25 (mean: 9.49). And the gamma richness ranged from 1 to 59 (mean: 18.59).

For the functional trait measurement, we aimed to sample the most dominant species at each site that account for more than 80% of the total biomass of this site, following the recommendation of de Bello et al. (2021). Original Fig. S9 was intended to demonstrate that we met this criterion across all sites. In the original Fig. S9, we plotted the *proportion* of biomass represented by species with trait measurements relative to the total aboveground biomass at each site—not absolute quantities. However, we agree that the figure could be difficult to read, particularly due to the display of all individual site IDs. To improve clarity, we have revised Fig. S5 (previous Fig. S9) to show a density histogram of these biomass proportions across all sites, rather than individual site-level bars. We hope this updated figure addresses your concerns and better conveys our sampling effort.

However, we felt that including species richness in this figure would not add meaningful information, as the focus is on trait sampling coverage in terms of biomass. For example, in some plots, species dominance was highly uneven, with a single species contributing the vast majority of the community biomass. In such cases, sampling that single dominant species was sufficient to represent most of the community biomass for trait measurements.

Line 297: "randomly-chosen" does not need to be hyphenated

> Response #3-13: we have used "randomly chosen" instead.

Lines 297-299: Did the authors take any measure to prevent water loss (and therefore leaf area reduction) between the sampling and the scanning? (e.g., keeping the samples in hermetic cold environments). If not, how could the area reduction by water loss have affected the results?

> Response #3-14: we appreciate the reviewer's thoughtful question regarding potential water loss during leaf sampling and its impact on trait measurements. To minimize water loss and preserve leaf lamina, we followed several precautionary steps. Leaves were not detached from the plants in the field until just before scanning. Before that, individual plants were temporarily stored in moist paper envelopes, which were then placed in sealed plastic bags to retain humidity. All samples were kept in a cooler with ice packs immediately after collection and transported to the lab, where scanning was conducted within ~12 hours in the same day. These procedures, consistent with standard protocols (e.g., Cornelissen et al. 2003), helped ensure that leaf area measurements were not significantly affected by dehydration. We have added these details to the revised Methods section in L316-324 for clarity.

Lines 296-307: The fate of the biomass sampled is unclear. Do the authors determine N and P from the very same leaves used to determine SLA? From a subsample of it? When describing SLA analyses, why do the authors reference the amount of biomass, and then, when describing N and P analyses, why do they refer to the fraction of species sampled? Did the authors take two different subsamples from each sample to determine SLA and nutrient contents? If so, is there any systematic difference between the subsamples taken for each trait? What effect could it have on the analyses? Please clarify.

> Response #3-15: thank you for these important methodological questions. For the aboveground biomass we sampled in each subplot, we sorted it by species and dried at 65 °C to a constant weight. We have added these details in the Methods section. In this study, the biomass data was mainly used to (1) calculate CWM and RaoQ by using the relative biomass of each species in the community; (2) test the relationship between biomass and NDVI to prove that NDVI can be a proxy as aboveground productivity; (3) prove that the accumulative biomass of the species we sampled for functional traits cover most of the biomass of a community (> 80%).

All leaf trait measurements (SLA, N, and P content) were conducted using the same set of

leaves sampled at each site. Immediately after scanning for SLA determination, the identical leaves were oven-dried and subsequently used for nutrient analyses. For both SLA and lead N and P, when we mentioned the fraction of species sampled, we wanted to prove that we have considered dominant biomass-contributing species in the communes, which meet the 80% criterion recommended by de Bello et al. (2021). We have revised the corresponding sentences to avoid confusion. The biomass coverage for nitrogen and phosphorus measurements (81.5%) was slightly lower than that for SLA (89.8%), primarily because the leaf material available from some species was insufficient for nutrient analysis, which requires a minimum sample quantity.

We did not create separate subsamples for different traits, thus eliminating potential subsampling biases. This consistency guarantees direct comparability between SLA and nutrient traits. We have clarified these methodological details in the revised Methods section.

Lines 296-307: The authors used NDVI as a proxy for vegetation productivity, however. While this index has traditionally been used, recent studies have reported better proxies, in particular, when vegetation cover is low, such as the near-infrared reflectance of vegetation (NIRv ~ NDIV * R_NIR, Badgley, et al. 2017), and even better when the incoming radiation is taken into account (Dechant et al., Dechant, et al. 2022). Have the authors considered using better estimates of vegetation productivity, such as NIRvP? While NIRv can only approach vegetation productivity at monthly/seasonal scales, Dechant et al. show that for shorter scales, the role of incoming radiation is relevant. On the one hand, these estimates could improve the characterization of the stability across sites featuring different illumination conditions; on the other hand, this work could also provide insight into how much ecological studies' conclusions are biased by using NDVI. The near-infrared reflectance (R_NIR) could be extracted from GEE, and incoming PAR could be approximated with the FLDAS Surface downward shortwave radiation (SWdown_f_tavg * 0.45).

> Response #3-16: we followed this suggestion and calculated the NIRvP index. The methods for the calculation are as below:

First, we extracted near-infrared reflectance (NIR) from Landsat 7 ETM+ (SR_B4) and Landsat 8 OLI-2 (SR_B5) surface reflectance products for each subplot from 2013 to 2022 based on their geographic coordinates (latitude/longitude) using the Google Earth Engine (GEE) platform (30-m resolution). Then we calculated the NIRv index (Near-Infrared Reflectance of Vegetation) as:

$$NIRv = NDVI \times NIR$$

Where NDVI (the normalized difference vegetation index) was obtained exactly the same as we described in the main text. In order to compare the previous results based on NDVI, we followed the same procedure as before and adopted the method of Oehri et al. (2017) to smooth NIRv time series. The average NIRv in growing season start (SOS) and end (EOS) time

span was estimated and used as the proxy of primary productivity of the year (the estimates from the densely smoothed curve were used to calculate the average value).

Next, we extracted surface downward shortwave radiation (SW_{down}) from FLDAS for each site from 2013 to 2022 based on their geographic coordinates. As the spatial resolution of FLDAS is 0.1 × 0.1 degrees, the four subplots at each site have the same SW_{down} value. Then we calculated NIRvP as:

$$NIRvP = NIRv \times PAR_{in}$$

Where $PAR_{in} = SW_{down} \times 0.45$, so

$$NIRvP = NDVI \times NIR \times SW_{down} \times 0.45$$

Similar to the approach used for NDVI in Fig. S6, we examined the relationship between NIRvP for each subplot and the aboveground biomass measured in the field in 2021 (or 2022), using both (c) the original data and (d) log-transformed data. We found that the R^2 values from linear models using NDVI were consistently higher than those using NIRvP for both regions, across both data types. These results suggest that NDVI is a more reliable proxy for aboveground plant productivity than NIRvP.

We also calculated gamma stability, alpha stability, and spatial asynchrony using NIRvP and constructed a multigroup piecewise structural equation model (SEM). The results (Fig. S14-15) showed similar patterns to those based on NDVI (Fig. 3-4), and the overall conclusions remained unchanged. Therefore, we kept the results based on NDVI in the main text. We added this information related to NIRvP in Supplementary Methods 1.

Lines 351-365: The use of the different weights is unclear. The authors use "unweighted relative abundance" to calculate Simpson diversity but "weighted" averages of community stability across subplots. How was relative abundance calculated using the number of individuals, the leaf area, the biomass, other…? What are the weights used in the second case, and why do weights differ between metrics? Please clarify.

> Response #3-17: thank you for your valuable comment. We calculated species relative abundances using the "unweighted" form, where each species contributes equally. In a community, w_c (a weighting parameter for the c_{th} species) was calculated as 1/n, with n being the number of the species. We used the "unweighted" because in general, w_c should be equal to 1/n (de Bello et al. 2010). We added more details in L412 as Response #3-18. Since Simpson diversity is strongly correlated with species richness, we have removed it from the updated version for simplicity.

For community stability, we have revised the definition as "the weighted average of community stability across subplots, weighted by the mean total biomass of each subplot". We calculated gamma stability, alpha stability and spatial asynchrony following the method

from Wang et al. (2019). Local-scale community variability (or alpha invariability) reflects the relative variability of community biomass at the local scale, quantifying how much biomass fluctuates in relation to its mean (Wang et al. 2019). Please, see also our response to Reviewer 1 (Response #1-5).

Lines 375-377: How were the community means weighted, using the number of individuals, the leaf area, the biomass, other…? Please clarify. The same applies for the "unweighted relative abundance" in line 378. Specify if this is the same used for Simpson, and if calculated differently, indicate also how it was calculated.

> Response #3-18: community weighted mean was calculated as the average of species trait values weighted by their relative biomass in the community, which ensures that dominant species contributing more to ecosystem functioning are appropriately weighted. We used "unweighted" relative abundance, where each species contributes equally, as explained above (Response #3-17). We have added these details in the text.

Line 399: Check the "or", should it be "on", or "or between"? The statement is unclear.

> Response #3-19: we have revised the Statistical analyses based on our updated statistics. And this sentence has been removed in the updated version.

Line 402: "(Fig. 13)", I think the authors mean Fig. S14.

> Response #3-20: yes. However, we have replaced it with the Moran's I test for spatial autocorrelation in the revised version.

Line 406-409: If the predicted variables (alpha stability and spatial asynchrony) were added to Fig. S14, the figure would also give an idea of how predictors correlate with them.

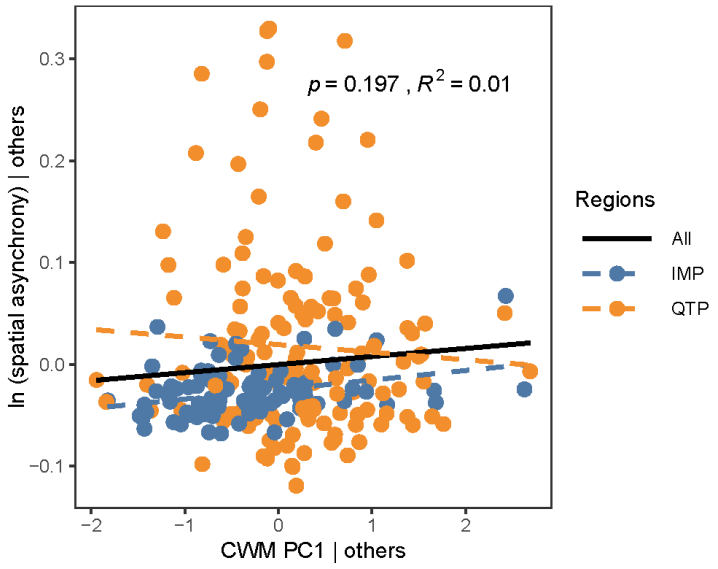
> Response #3-21: we have added alpha stability and spatial asynchrony to Fig. S9 (originally Fig. S14). The results showed that neither alpha stability nor spatial asynchrony is strongly correlated with the predictors.

Results
Figure 3: Figures should be self-explanatory. Please add to the caption what "richness", "Regions", and "richness:Regions" stand for, and as in Figure 5 caption, define the symbology associated with the degree of significance. Do the same for Figure 4.

> Response #3-22: we have removed the previous Figs 3-5 in the revised version. To ensure clarity, we checked that all figure captions are self-explanatory.

Lines 114-118: Could the fast-slow functional composition explain the spatial asynchrony? Has this been tested?

> Response #3-23: in theory, the fast-slow functional composition reflects the dominance of the species with traits related to fast or slow resource-use strategies, while spatial asynchrony is the asynchronous dynamics of productivity across local communities. To our knowledge, there is no established theoretical framework or hypothesis that directly links these two aspects. Nevertheless, to address this question, we estimated the partial effects of CWM PC1 (the first principal component of community weighted mean of SLA, leaf N and P content) on spatial asynchrony. The result shows no significant relationship of CWM PC1 and spatial asynchrony after controlling the effects of other variables.



F

Line 112: Define "iv" the first time it is used. It is currently defined in line 414

> Response #3-24: we have removed the model selection analyses following the suggestions from Reviewer 1 (Response #1-6), as well as all the sentences related to importance value (iv) in the revised version.

Discussion

Lines 191-193: Does the referenced study also focus on grasslands? Specify.

> Response #3-25: we have verified that the study cited in L183 focuses on grassland ecosystems. We have added this detail in the text.

Lines 197-212, 221-223: Regarding the recommendation of using NIRv (see methods), notice that NDVI can induce uncertainties in the estimation of productivity when there is bare soil in the pixel. NIRv (and other indices) are more robust to that. Understanding whether NDVI is a good enough proxy to assess the effect of disturbance/aridity on productivity would be interesting.

> Response #3-26: thanks for the reviewer's valuable suggestion. We fully acknowledge the limitations of NDVI in areas with bare soil exposure and appreciate your recommendation to consider NIRvP as a more robust alternative. In response, we have conducted parallel analyses using both NDVI and NIRvP indices, with the results of NIRvP presented in Fig. S14-15. These results were highly consistent with those based on NDVI. Therefore, we kept the results based on NDVI in the main text, and provided the NIRvP-based results in the Supplementary Information, with more details responded above (Response #3-16).

In addition, NDVI allows for direct comparison with previous studies, many of which also relied on NDVI. For example, García-Palacios et al. (2018) demonstrated that NDVI correlates well with productivity in dryland ecosystems, despite its known limitations.

Lines 234-242: This might result from not accounting for incoming radiation when estimating productivity and its stability. NIRvP-based stability estimates might provide different results (and maybe more consistent with the literature).

> Response #3-27: we have conducted comparative analyses using both conventional NDVI and the radiation-corrected NIRvP approaches. Our results show relatively consistent patterns between NIRv and NDVI estimations.

Regarding the neutral relationship between beta diversity and spatial asynchrony, we have added more potential explanations in the Discussion according to the comments from Reviewer 1 (Response #1-4) and 2 (Response #2-5).

Line 270: This statement needs to be corrected: "linked to alpha stability at local, alpha spatial scales"

> Response #3-28: we have revised it as "contribute to alpha stability".

Reviewer #5 (Remarks on code availability):

I did not run the code and R is not the programming language I know the best. However, the code seems sufficiently commented and includes human-readable .csv files, a README.txt file describing the other files content and role. The necessary packages are installed when R code is executed.

> Response #3-29: we appreciate your effort in reviewing the code. We have strived to make the R scripts as clear and accessible as possible by providing more detailed comments.

References:

Bai, Y., Han, X., Wu, J., Chen, Z., & Li, L. (2004). Ecosystem stability and compensatory effects in the Inner Mongolia grassland. *Nature*, *431*(7005), 181-184.

Bullock, J. M., Mallada González, L., Tamme, R., Götzenberger, L., White, S. M., Pärtel, M., & Hooftman, D. A. (2017). A synthesis of empirical plant dispersal kernels. Journal of Ecology, 105(1), 6-19.

Cornelissen, J. H., Lavorel, S., Garnier, E., Díaz, S., Buchmann, N., Gurvich, D. E., ... & Poorter, H. (2003). A handbook of protocols for standardised and easy measurement of plant functional traits worldwide. Australian journal of Botany, 51(4), 335-380.

Carmona, C. P., Bueno, C. G., Toussaint, A., Träger, S., Díaz, S., Moora, M., ... & Tamme, R. (2021). Fine-root traits in the global spectrum of plant form and function. Nature, 597(7878), 683-687.

Chen, L., Jiang, L., Jing, X., Wang, J., Shi, Y., Chu, H., & He, J. S. (2021). Above-and belowground biodiversity jointly drive ecosystem stability in natural alpine grasslands on the Tibetan Plateau. Global Ecology and Biogeography, 30(7), 1418-1429.

de Bello, F., Carmona, C. P., Dias, A. T., Götzenberger, L., Moretti, M., & Berg, M. P. (2021). Handbook of trait-based ecology: from theory to R tools. Cambridge University Press.

De Bello, F., Lavergne, S., Meynard, C. N., Lepš, J., & Thuiller, W. (2010). The partitioning of diversity: showing Theseus a way out of the labyrinth. Journal of Vegetation Science, 21(5), 992-1000.

ECVC (The editorial committee of vegetation of China) (1980). Vegetation of China (in Chinese). Beijing: Science Press.

García-Palacios, P., Gross, N., Gaitán, J., & Maestre, F. T. (2018). Climate mediates the biodiversity–ecosystem stability relationship globally. Proceedings of the National Academy of Sciences, 115(33), 8400-8405.

Hautier, Y., Zhang, P., Loreau, M., Wilcox, K. R., Seabloom, E. W., Borer, E. T., ... & Wang, S. (2020). General destabilizing effects of eutrophication on grassland productivity at multiple spatial scales. Nature Communications, 11(1), 5375.

Lososová, Z., Axmanová, I., Chytrý, M., Midolo, G., Abdulhak, S., Karger, D. N., ... & Thuiller, W. (2023). Seed dispersal distance classes and dispersal modes for the European flora. *Global Ecology and Biogeography*, *32*(9), 1485-1494.

Ma, W., He, J. S., Yang, Y., Wang, X., Liang, C., Anwar, M., ... & Schmid, B. (2010). Environmental factors covary with plant diversity–productivity relationships among Chinese grassland sites. *Global Ecology and Biogeography*, *19*(2), 233-243.

Ma, Z., Liu, H., Mi, Z., Zhang, Z., Wang, Y., Xu, W., ... & He, J. S. (2017). Climate warming reduces the temporal stability of plant community biomass production. *Nature Communications*, *8*(1), 15378.

Oehri, J., Schmid, B., Schaepman-Strub, G., & Niklaus, P. A. (2017). Biodiversity promotes primary productivity and growing season lengthening at the landscape scale. Proceedings of the National Academy of Sciences, 114(38), 10160-10165.

Sun, H. L. (2005). China Ecosystem (in Chinese). Beijing: Science Press.

Wang, C., Wang, X., Liu, D., Wu, H., Lü, X., Fang, Y., ... & Bai, E. (2014). Aridity threshold in controlling ecosystem nitrogen cycling in arid and semi-arid grasslands. Nature communications, 5(1), 4799.

Wang, S., Lamy, T., Hallett, L. M., & Loreau, M. (2019). Stability and synchrony across ecological hierarchies in heterogeneous metacommunities: linking theory to data. Ecography, 42(6), 1200-1211.

Wang, Y., Niu, X., Zhao, L., Liang, C., Miao, B., Zhang, Q., ... & Ma, W. (2020). Biotic stability mechanisms in Inner Mongolian grassland. *Proceedings of the Royal Society B*, *287*(1928), 20200675.

Whittaker, R. H. Evolution and measurement of species diversity. Taxon 21, 213-251 (1972).

REVIEWER COMMENTS

Reviewer #1 (Remarks to the Author):

Thank you for this opportunity to review the revisions to this manuscript. I have reviewed all the authors changes, and I am pleased that most of my earlier concerns have been addressed. I appreciated the caveats that the authors added to the discussion regarding spatial scale and unmeasured traits. I still think Figure 1 needs work to reach its potential, and provide some potential solutions below.

> Response #1-1: We are grateful to the reviewer for these positive comments, and appreciate the suggestions for Figure 1; we have followed them to further improve it (see below). Line numbers refer to the revised version with tracked changes.

Abstract: please change "contributes to (de)stabilize" to "contributes to (de)stabilizing" throughout, to be grammatically correct.

> Response #1-2: Agreed. We have changed this accordingly.

Results: In "…and unrelated to the dominance of fast-growing species, as indicated by CWM PC1." I suggest using "summarized" rather than "indicated"

> Response #1-3: We have implemented the reviewer's suggestion.

Discussion: I had trouble understanding in this sentence if you were highlighting a difference between the SEM plots and the partial regression plots, or something else: "Although higher mean annual precipitation and temperature 360 favored dominance by fast-growing species, this did not translate into higher alpha stability; a pattern that contrasts with results from partial regression plots."

> Response #1-4: We apologize for the lack of clarity. With this sentence, we wanted to highlight the difference between the SEM results and the partial regression results: CWM PC1 was not related to alpha stability in SEM, but positively related to alpha stability in the partial regression results. We have revised the sentence to make this distinction clearer in L291-293: "…this did not translate into higher alpha stability, in contrast to the partial regression plots where fast-growing species dominance was positively linked to alpha stability…".

Figure 1. The figure is still not understandable without a very detailed reading of the legend

(which while much more useful, has now become very complicated). Part of the issue with the figure itself is that the SEM uses terms (e.g. CWM PC1, apha RaoQ, beta RaoQ) that are not used in the panels a-d so it very difficult for the reader to relate the different panels of the figure. Why not put on the "trait value" axis the words "slow" and "fast" on the left and right extremes of the axis to indicate the slow to fast strategy trait gradient and modify "trait value" to be "trait value (PC1)"? It seems like the major difference between (a)&(c) versus (b)&(d), is whether functional divergence as measured by alpha RaoQ is low or high, and whether functional diversity among communities as measured by beta RaoQ is low or high. I note that you have carefully drawn each row of panels so the same two communities are shown, but different aspects of these communities are highlighted in different panels on the same row. Consider therefore adding vertical blue panel strips (similar to the horizontal blue panel strips) that say "Low alpha RaoQ, high beta RaoQ " and "High alpha RaoQ, low beta RaoQ". If instead you prefer to label (a),(b), (c), (d) individually then labels such as "low alpha RaoQ", "high alpha RaoQ", "high beta RaoQ", "low beta RaoQ" could be used, but then you really don't need the second community in plots (a) and (c) because this is really a temporal comparison of a community to itself.

> Response #1-5: We appreciate the reviewer's suggestions. We have followed them and indicated the words "slow" and "fast" on the trait value axis to make the slow to fast strategy trait gradient more obvious. However, we have kept the label "trait value" unchanged, because PC1 was only used for CWM, but not for RaoQ in our analyses.

We have also added vertical blue panel strips with "Low alpha RaoQ, high beta RaoQ" for panels (a) and (b), and with "High alpha RaoQ, low beta RaoQ" for panels (c) and (d).

Figure 3. Please, in the legend, tell the reader where they can learn the identity of "others" (e.g. a supplementary table, etc). I am also concerned that the authors are showing regression slopes for non-significant relationships. The convention is that regression slopes are only shown when they are significant, otherwise just the cloud of points should be shown.

> Response #1-6: Following the reviewer's recommendation, we have added more details about "others" in the figure legend, and also added Table S3 to show which variables are exactly included in "others". We have also revised Fig. 3, Fig. S4, and Fig. S14 to show the regression lines only if the regression across all sites is significant ($p < 0.05$).

Figure 4. Please explain the cut-off for significance of paths, as there seems to be a discrepancy in the legend "Non-significant paths are not shown ($P \geq 0.1$). Numbers on arrows are standardized path coefficients (scaled by their mean and standard deviation), and asterisks indicate statistical significance (***$p < 0.001$; **$p < 0.01$; *$p < 0.05$)." My guess is that the

authors originally showed marginally non-significant paths (0.1>p>0.05), but now do not and forget to adjust the first sentence.

> Response #1-7: We thank the reviewer for noticing this detail. Indeed, we originally showed marginally non-significant paths ($0.05 < p < 0.1$), then later removed them from the figures and the results for simplicity. We have revised the sentence in the legend in L724-727 as: "Black arrows represent significant constrained paths ($p < 0.05$)… Non-significant paths are not shown ($p \geq 0.05$)".

Reviewer #1 (Remarks on code availability):

Addressed in first review, still acceptable.

> Response #1-8: We are grateful to the reviewer for the positive feedback on the code.

Reviewer #2 (Remarks to the Author):

Review
The authors have well addressed the comments raised in the previous review. In line with the recommendations, they have now (i) strengthened the theoretical foundation of the framework they are testing, (ii) clarified their methods and statistical analyses, and (iii) streamlined their analytical approach (for instance, by removing the model selection analysis). Overall, I think that these revisions have considerably improved the manuscript. However, I believe that several additional changes would further enhance the clarity and impact of the paper. More specifically:
1. In the discussion, the authors could further elaborate on the underlying mechanisms and processes at play. In particular, they could revisit the components of stability (resilience and resistance), which are now mentioned in the introduction, and link these mechanisms more explicitly to their results to provide a stronger interpretation.

> Response #2-1: We thank the reviewer for this comment. Given that this point is discussed more in detail below, we provide an elaborate answer in Response #2-9.

2. The authors should ensure that the new sections added to the discussion remain fully coherent with the rest of the manuscript. I recommend a careful re-reading to strengthen the logical connections between ideas and to make the arguments more explicit.

> Response #2-2: We agree with the reviewer that a more cohesive text is needed. We have now revised and strengthened the logical connections between ideas in the Discussion. Also see our response in Response #2-14.

Specific comments (line numbers refer to the revised version with tracked changes)

• L34: Why was the number of plots changed in this revised manuscript?

> Response #2-3: In previous versions of the manuscript, we excluded seven sites because there was insufficient spatial distance between them, thus they occupied the same pixel and had the same NDVI value. However, during our latest re-analysis, we identified two more site that also met that criterion for exclusion, making it a total of nine excluded sites. We clarify this decision in L378-380.

• L40: Should this not read "relationships between environmental conditions and local diversity and stability"?

> Response #2-4: Absolutely. We have now reworded this sentence as suggested.

• L43–46: The sentence "Our study offers new insights into how functional traits along the fast–slow leaf economics spectrum mediate the effects of environmental change factors (e.g., precipitation) on ecosystem stability at contrasting spatial scales" feels somewhat disconnected from the preceding part of the abstract. If you wish to highlight the fast–slow spectrum here, you should also mention results on the relationship between these traits and stability in the abstract I think.

> Response #2-5: We thank the reviewer for their suggestion. In our study, we used functional traits along the fast–slow leaf economics spectrum (specifically CWM and RaoQ) together with SEM to examine how environmental factors influence ecosystem stability through their effects on CWM and RaoQ. Among these, CWM was not significantly related to stability, while RaoQ was negatively correlated with alpha stability (as reported in the preceding part of the abstract). Therefore, we believe that we do focus on the relationship between the fast–slow spectrum and stability, and we have clarified the nature of these relationships in the abstract.

• L109: You also examine the relationships between species richness, functional diversity, and composition, since you consider the effects of CWM as well. Please clarify.

> Response #2-6: We revised "functional diversity" as "functional diversity and composition" to also include CWM.

• L126: "CWM PC1" – this variable should be introduced more clearly. As results appear before the methods, a brief explanation here of what is exactly this variable would help the

reader.

> Response #2-7: We have clarified that CWM PC1 represents the first axis of a principal component analysis of the community weighted mean of fast strategy traits. We have provided a similar explanation for RaoQ as well.

• L196: From a semantic perspective, I would say that it is preferable to avoid saying that an effect "disappears". Instead, write that the effect "was not observed"?

> Response #2-8: Agreed. We have modified this as suggested.

• L212–214: Please expand here on the underlying mechanisms. For instance, describe how fast strategies may enhance stability and link back to the mechanisms outlined in the introduction (e.g., improved resilience).

> Response #2-9: We thank the reviewer for this comment. In the Introduction, we hypothesized that "communities dominated by fast species can exhibit high community stability due to high resilience, whereas communities dominated by slow species can also exhibit high community stability due to high resistance" (L57–60). Based on this framework, we expected that resilience or recovery would be linked to dominant species, as measured by CWM. However, our results showed that functional diversity (measured by alpha RaoQ) was negatively related to alpha stability, while CWM was not related to alpha stability. We interpreted this pattern in light of the second mechanism outlined in the Introduction, namely negative covariance arising from competition, which we discuss in L295–300. Furthermore, we also addressed the relationship between CWM of fast-strategy traits and alpha stability, and the potential links to resistance and recovery, in L191–205.

• L294: Would "potentially greater" not be more accurate here?

> Response #2-10: Yes, we agree. We have replaced this expression.

• L302: Typo: should read patterns.

> Response #2-11: Revised.

• L310: Replace "disappeared" with "was not observed"

> Response #2-12: Done.

• L314–320: Would you really expect that considering belowground traits would result in a stronger effect of beta FD on stability? While the discussion of dispersal traits is interesting in this context, I wonder if the section on belowground traits is appropriate here when discussing beta stability.

> Response #2-13: Indeed, we expect that belowground traits may lead to a stronger relationship between beta functional diversity and stability. For instance, precipitation is an important environmental factor in our study. Compared to aboveground traits such as SLA, the response rate and extent of belowground traits such as root depth to drought can be very different. This response diversity across heterogeneous environments can enhance spatial asynchrony among communities, thereby strengthening the relationship between beta functional diversity and stability.

• L321–334: This section is unclear. I would consider moving this information to the Methods section instead. I understand it was added in response to reviewer comments, but it seems more appropriate for the Methods. Again, ensure that the additions to the discussion remain coherent with the overall flow of the paper.

> Response #2-14: This paragraph is our third and final point to discuss the potential effect of our experiment design on the neutral relationship between beta richness or functional diversity and spatial asynchrony. It is the continuation of the previous paragraph, but from the perspective of plot design. We think such a reflection on experimental design is better suited for the Discussion than Methods, since it is not vital to understand the design but it conveys important messages that we want the readers to bear in mind while considering the results of this manuscript. That said, we truly appreciate the reviewer's comment and agree that there is room for improvement, thus we have revised this paragraph to ensure its coherence with the overall flow of the paper.

• L329: The notion of "mismatches" is unclear. I could follow because I had read the review responses, but in the manuscript itself this is difficult to understand. Could you please clarify that section? Again, should it not be presented in the Methods section instead?

> Response #2-15: We have made this sentence clearer in L273-275: "…the scale differences between field and remote sensing data…". Regarding the placement, we believe that it is better suited for the Discussion, as reasoned in Response #2-15.

• L360: Replace "dominance by fast-growing species" with "the dominance of fast-growing species."

> Response #2-16: Replaced as suggested.

• L375: Please expand on the implications of your results. For instance, in the conclusion you write: “In conclusion, our results demonstrate that the functional traits of dominant species influence ecosystem stability. While species richness and functional diversity contribute to alpha stability at local scales, they show no relationship with spatial asynchrony at larger scales. The influence of environmental conditions on diversity and alpha stability differs between regions, yet the relationship between alpha diversity and stability is universally consistent.” This is an important conclusion, but its broader ecological or management implications could be made clearer.

> Response #2-17: We appreciate this suggestion and have now added a sentence to show the broader ecological or management implications: “It suggests that management strategies aimed at maintaining local diversity can reliably support ecosystem stability across different environmental contexts” (L305-306).

• L406: Is the 10.2% referring to overlap in species composition or in the species pool? Please specify.

> Response #2-18: We apologize for the lack of clarity. The 10.2% refers to the proportion of species shared between the two regions relative to the total species pool. To avoid confusion, we have revised this sentence in L332-333.

• L774: Add “positively” before “related” for precision.

> Response #2-19: Added.

• L798–801: Please clarify that the paths from abiotic factors to alpha stability and spatial asynchrony are not shown but are nevertheless included in the SEM.

> Response #2-20: We appreciate the suggestion and have added this information accordingly.

• Figure 1: Why not include a covariance path between CWM PC1 and Alpha RaoQ? Based on expectations, plant communities dominated by fast-growing species should be less functionally diverse no?

> Response #2-21: We thank the reviewer for their question. We included the covariance path between CWM PC1 and alpha RaoQ in the SEM model. We also included other strong

correlated errors (e.g., X %~~% Y) based on the directed separation (d-sep) approach implied by the initial model to improve model fit (L457-460; details available in the R code). We then applied a multi-group SEM to test whether the pathways differ between the two regions. However, the covariance paths are not displayed in the output of the "multigroup" function, and we also did not show them in the a priori SEM model for clarity. We have now added this detail to the figure legend.

• Figure 1e: You could use a more self-explanatory label for "CWM PC1"

> Response #2-22: We thank the reviewer for their suggestion. While a more self-explanatory label would be advisable in many contexts, here we prefer to prioritize keeping Fig. 1e consistent with other figures (Figs. 3–4; and particularly Fig. 4c), where abbreviations are usually necessary due to limited space. Nevertheless, to ensure that the reader has all the information they need to understand all labels, we have clearly defined "CWM PC1" in all figure legends.

• Figure 4: In panels (a) and (b), only significant paths are shown, whereas in panel (c) both significant and non-significant paths are displayed. Why this inconsistency? It would be clearer to adopt the same approach across panels.

> Response #2-23: Panels (a) and (b) present the results of the multi-group SEM, highlighting how the pathways differ between the two regions, with significance explicitly tested. In contrast, panel (c) displays the direct, indirect, and total effects of variables on alpha stability and spatial asynchrony for both regions, calculated from the standardized path coefficients of the SEM, without emphasizing significance. These two types of results complement each other by illustrating different aspects of the model but treat significance differently; therefore, the use of paths is not inconsistent but representative of the analysis. Despite of this, we completely understand the potential confusion and we have revised the caption to clarify the different representations.

Reviewer #2 (Remarks on code availability):

The analyses and results of the paper are reproducible and the code can be easily used by the community. A 'README' file is provided and it contains instructions in order to reproduce some of the analyses, figures and results.

> Response #2-24: We thank the reviewer for their thoughtful comments on our code.

Reviewer #5 (Remarks to the Author):

GENERAL COMMENTS

The authors have addressed the questions and suggestions made in the first review. Following other reviewers' recommendations, SEM analyses have been simplified, and additional checks have been applied. I thank the authors for testing the use of NIRvP; the comparison is valuable for future analyses, even if eventually this index did not yield improved results. The updated version of the manuscript reads more clearly after key statements were rewritten or corrected.

I just have a few suggestions to clarify some aspects that remain confusing

> Response #3-1: We thank the reviewer for their positive feedback on the last revision of our manuscript, and have followed their suggestions to further improve the manuscript.

SPECIFIC COMMENTS (lines refer to the version with tracked changes)

Line 31-33: From the statement, it seems that the reason why it is unknown that the relationships hold at larger spatial scales is management; however, the study avoids managed areas. It rather seems that the reason behind this is an open question is it has not been addressed, as later suggested in the introduction. Maybe management is not so relevant in this study as to include it in the abstract.

> Response #3-2: We thank the reviewer for pointing this out. We agree that mentioning management here may be misleading, as our study avoids managed areas. We have revised this sentence to clarify it as: "It remains unclear whether these relationships persist at larger spatial scales".

Lines 265-267: "suggesting that species in the desert steppe tend to aggregate in patches with relative low plant cover due to strong abiotic heterogeneity." It is unclear where this conclusion comes from. Is it taken from the reference 47 (Sasaki, et al. 2024) or is it made by the authors? This statement requires further explanation and references that support it or a clearer link to the work of Sasaki.

> Response #3-3: We have now revised this sentence, and provide a clearer explanation of the preceding statement from the perspective of mechanisms related to diversity-stability relationship that were supported by Sasaki et al. (2024), in L214-217.

Lines 302-304: If homogeneity can explain the lack of relationship between beta diversity and stability at larger scales, could the steppe sites with high abiotic heterogeneity be used to test whether this hypothesis holds?

> Response #3-4: We thank the reviewer for the question. We would expect that, at steppe sites such as desserts, abiotic heterogeneity was high at the microhabitat scale (i.e., within communities). In contract, the abiotic conditions are likely to be relatively homogeneous among communities. We also tested the relationships between beta richness, beta RaoQ and spatial asynchrony across different vegetation types (Fig. S2d-e). The results showed that the relationship between beta richness and spatial asynchrony was in fact slightly negative, which does not support our hypothesis.

Line 362: Point to the figure subplot that the statement refers to.

> Response #3-5: Indeed. We have followed this suggestion.

Line 404: The samples were "dried at 65 °C to a constant weight"; how was this achieved? Were the samples regularly measured or just dried for a sufficiently long time? If so, what was the drying time?

> Response #3-6: We dried the biomass samples in the oven at 65 °C for a sufficiently long period to ensure constant weight. Based on our previous experience with similar material, this typically required 72 hours. We have revised this sentence in L344-345.

Lines 406-424: The discrepancy between the percentage of biomass that the species used for SLA (89.8%) and NP determination (81.5%) remains unclear. The authors indicate in the answer to the reviewer comments that not all the SLA samples could be used for NP determination due to insufficient sample quantity. I guess this means that some species were not included in the NP analyses, likely due to small leaf sizes.
However, the differences in the shape of the histograms in Fig. S5 make me wonder whether the specie/s not sampled for NP were present in a large number of plots. Also, once I understood that SLA and NP come from the same leaves but that NP analyses were not performed in some cases due to insufficient samples, I wonder how the original Fig. S9 shows a percentage of species biomass lower for SLA than for NP in some cases. I understand some data might have been filtered after determining SLA due to errors or large uncertainty.
How were samples where SLA or NP values were missing handled to compute plant functional diversity? Was there any sort of gap-filling? This all should be clarified in the methods, or if too detailed, in a supplementary section.

> Response #3-7: We truly appreciate these detailed comments. We have now carefully double checked our data and clarified the procedures. Our aim was to measure both SLA and NP of the same species at each site, to avoid potential biases. However, there were some inevitable mismatches:

1. At some sites, certain species have NP data but no SLA data, which explains why the percentage of biomass covered by SLA is lower than by NP in a few cases (as in the original Fig. S9).
2. Among our 235 sites, 13 sites contain species with both SLA and NP data, 156 sites contain species with SLA data but at least one species lacking NP, and 64 sites contain at least one species with NP data but lacking SLA.

This discrepancy arose from our field and lab protocols. In the field, we selected individuals for SLA based on a visual estimation of dominant species, without prior data analysis. SLA was measured immediately on fresh leaves. Later, NP was determined from dried material (both the biomass samples and the leaves used for SLA). If the sample amount was insufficient, SLA could be available but NP missing. Conversely, in some cases only a few species were chosen for SLA, while more dominant species were included for NP analysis based on the dried biomass samples, resulting in sites where at least one species had NP but no SLA.

For the calculation of functional composition and diversity (CWM and RaoQ), we did not apply any gap-filling. Instead, we used the mean trait value of each species across sites, and excluded species with missing SLA or NP when calculating CWM PC1 and RaoQ. Our site-level trait sampling aimed (1) to minimize environmental bias in trait values, and (2) to enable future analyses of trait variability along environmental gradients. The test of biomass coverage ensured that the species with measured traits represented, on average > 80% of the total aboveground biomass at each site. We also tested the species biomass proportion at each site for species with complete SLA, N, and P data after calculating the mean trait value of each species across all sites, with an average of 94.3% (Figure S5c), even a higher proportion and still meet the 80% criterion.

We have revised the Methods (L347-348, L421-424) and Figure S5 to clarify.

Reviewer #5 (Remarks on code availability):

As in the first review, I did not run the code, as I cannot provide expert advice on programming language I know the best. However, the code seems sufficiently commented and includes human-readable .csv files, a README.txt file describing the other files content and role. The necessary packages are installed when R code is executed and the libraries referenced in the manuscript are included.

> Response #3-8: We thank the reviewer for their positive comments on the code.

REVIEWERS' COMMENTS

Reviewer #1 (Remarks to the Author):

Thank you for these revisions. They are all fine, expect I believe that the authors may have made a mistake in the lettering in the legend of Figure 1. The legend describes the effects of alpha RaoQ on alpha stability as occurring in (a) and (b), but my understanding is that on the actual figure these occur in (a) and (c). Similarly the effects of beta RaoQ on asynchrony are described in the legend as being in (c) and (d), but I see these in the figure itself on panel (b) and (d). My apologies if I've misunderstood - but in that case the legend really is not clear.

> Response #1-1: We thank the reviewer for noticing this important detail. We have corrected the legend, with (a) and (c) for the effects of alpha RaoQ on alpha stability, and (b) and (d) for the effects of beta RaoQ on spatial asynchrony.

Reviewer #1 (Remarks on code availability):

Code was reviewed on first instance and hasn't appreciably changed.
> Response #1-2: We thank the reviewer for the comments on our code.

Reviewer #2 (Remarks to the Author):

General comments

Although the authors have addressed some of the previous revisions, I would still strongly recommend to improve the logical flow of ideas in the discussion and to make the arguments more explicit. As it stands, I found that it remains somewhat superficial. The implications of the findings and the mechanisms involved could be examined in greater depth I think. I would recommend a revision to strengthen this section.
> Response #2-1: We thank the reviewer for this constructive comment. Following the reviewer's suggestion, we have revised the Discussion to improve the logical flow and to make the arguments and underlying mechanisms more explicit, as detailed in Response #2-6 below.

Specific comments (line numbers refer to the revised version with tracked changes)

L31: Consider to add "however" before "whether".

> Response #2-2: We have followed this suggestion and added 'However' before 'it remains

unclear whether…' in L30.

L117: Is this "a principal axis of the principal component analysis of the CWM of fast strategy traits", or rather an axis associated with these CWMs ? Please be more precise.

> Response #2-3: We agree that the original wording was imprecise. Here, CWM PC1 refers to the first principal component derived from a principal component analysis of community weighted mean of fast strategy traits (i.e. SLA, leaf nitrogen, and leaf phosphorus content), rather than an axis associated with these CWMs. We have revised the text to clarify this point in L114-115.

L117: Replace "divergence" with "diversity" for consistency with the rest of the manuscript, and define it explicitly as measured by RaoQ.

> Response #2-4: We have replaced 'divergence' with 'diversity' accordingly, and defined it explicitly in L117.

L202: Replace "disappeared" with "was not observed" (as requested in the previous review).

> Response #2-5: We have followed this suggestion.

L254–260: Please incorporate elements from your response to clarify the reasoning. As currently written, it remains unclear why and how belowground traits would lead to a stronger relationship between beta functional diversity and stability.
Reference to the authors' response:
"Indeed, we expect that belowground traits may lead to a stronger relationship between beta functional diversity and stability… This response diversity across heterogeneous environments can enhance spatial asynchrony among communities, thereby strengthening the relationship…"
Please integrate this explanation more directly into the revised text to make your argument explicit.

> Response #2-6: We thank the reviewer for this suggestion. We have integrated the explanation from our previous response into the text to explicitly clarify how belowground traits could lead to a stronger relationship between beta functional diversity and stability, in L254-258.

L286: Replace "disappeared" with "was not observed" (as mentioned previously).

> Response #2-7: We have followed this suggestion.

Reviewer #2 (Remarks on code availability):

The analyses and results of the paper are reproducible and the code can be easily used by the community. A 'README' file is provided and it contains instructions in order to reproduce some of the analyses, figures and results.

> Response #2-8: We thank the reviewer for the comments on our code.

Reviewer #5 (Remarks to the Author):

GENERAL COMMENTS

The authors have addressed all the questions and suggestions made in the second review. I found only minor corrections needed in Fig. 1, along with a remark that the authors could easily address in one or two sentences. Beyond that, I think the manuscript could be accepted for publication.

> Response #5-1: We thank the reviewer for the positive comments. We have addressed the remaining minor points, as described below.

SPECIFIC COMMENTS

Figure 1a-d. In the y-axis labels, replace “alpah” with “alpha”.

> Response #5-2: We have followed and revised Figure 1a-d.

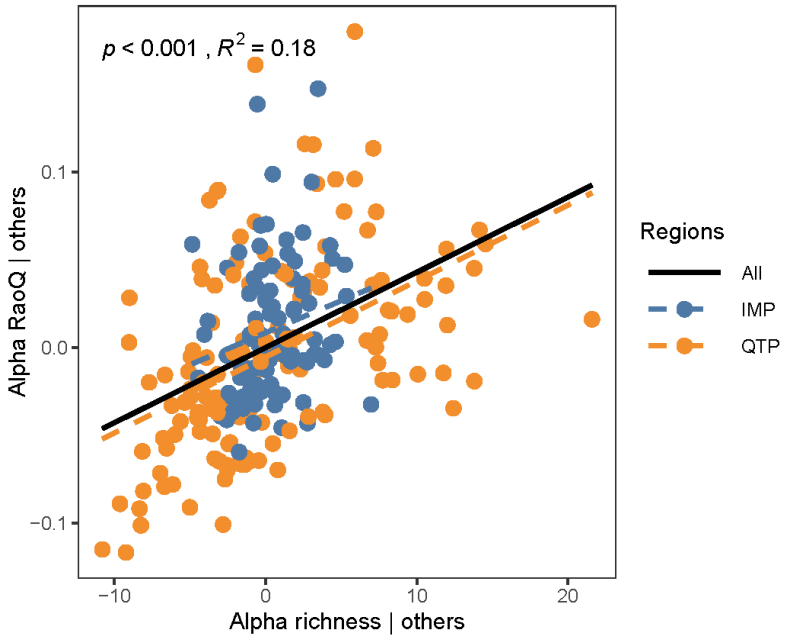
Figure 1 caption: Make sure the subplots are properly referenced; for example, subplots a and b are said to feature shaded areas representing individual species, but these appear in a and c.

> Response #5-3: We have revised the subplot reference, as responses in Response #1-1.

Lines 167-170: Is this result a consequence of a negative relationship between species richness and functional diversity related to the fast-slow leaf economic spectrum? Fig. 3a,c suggest so. If this is correct, please specify and indicate whether this correlation is expected or known.

> Response #5-4: We thank the reviewer for this question. From the SEM in Figure 4 and the Pearson correlation (correlation coefficient = 0.57) in Figure S9, we can see a positive, rather than a negative relationship between alpha richness and alpha functional diversity. We also tested the partial relationship between alpha richness and alpha functional diversity after removing the effects of all other variables. The result as below also shows a positive

relationship between alpha richness and alpha RaoQ. Therefore, the contrasting effects of species richness (stabilizing) and fast–slow functional diversity (destabilizing) on ecosystem stability cannot be attributed to a negative covariation between them. Instead, our results indicate that different dimensions of biodiversity influence ecosystem stability through distinct and partially independent mechanisms.



Reviewer #5 (Remarks on code availability):

As in the former reviews, I did not run the code, as I cannot provide expert advice on R programming language. However, the code seems sufficiently commented and includes human-readable .csv files, a README.txt file describing the other files content and role. The necessary packages are installed when R code is executed and the libraries referenced in the manuscript are included.

> Response #5-5: We thank the reviewer for the comments on our code.