Response to Reviewer Comment 1

Comment 1. The definitions of resistance and resilience need stronger conceptual and mathematical justification

We thank the reviewer for this important comment. We agree that the definitions of resistance and resilience are central to the interpretation of this study, and that the original manuscript did not sufficiently justify why we adopted these equations rather than more commonly used ratio-based, bounded, or log-response metrics.

We adopted this formulation because the objective of this study was to compare global, pixel-wise vegetation responses across land-cover types and hydroclimatic-event categories using a response variable with a direct and consistent interpretation. This formulation is centered around 1 and avoids some practical difficulties that may arise when applying alternative metrics to global gridded datasets. For example, ratio-based metrics such as the Isbell-type formulation can become extremely large or undefined when the event-year or post-event vegetation index is very close to the normal-year condition. Bounded metrics such as the Orwin and Wardle-type formulation are useful for quantifying departure from a control condition, but they treat above-normal vegetation activity as a deviation from the normal state rather than as enhanced vegetation activity. Log-response metrics are useful for quantifying proportional changes, but they are centered around 0 and do not directly indicate whether vegetation activity has returned to the normal-year condition.

To address the reviewer's concern more explicitly, we have added a new Supplementary Note and Supplementary Tables comparing our metrics with Isbell-type, Orwin and Wardle-type, and log-response metrics using the same hypothetical vegetation-index values(Table S1 and S2). These comparisons clarify how each metric behaves when vegetation activity remains close to the normal condition, declines below it, fully recovers, or exceeds the normal-year level. We also revised the manuscript text to state more clearly that our indices should be interpreted as proximity-to-normal vegetation activity indices, rather than as directly interchangeable with all previously proposed resistance and resilience metrics.

In addition, we have revised the Discussion to avoid overstating comparability with previous studies that used different formulations. We now explicitly acknowledge that different resistance and resilience metrics emphasize different ecological aspects of disturbance response, and that our results should be interpreted within the definition adopted here. We believe these revisions provide a clearer conceptual and mathematical justification for the indices used in this study and improve the transparency of the analysis.

Main manuscript: p. 8, lines 176–189. Supplementary Tables S1–S2.

Comment 2. NDVI as a choice for indicating ecosystem function

We thank the reviewer for this important suggestion. We agree that the choice of vegetation index is a critical issue in studies of ecosystem functioning and stability. We also agree that EVI has several advantages over NDVI, particularly because it can reduce saturation effects in high-biomass vegetation and includes corrections for soil and atmospheric background effects.

In this study, we used the GIMMS NDVI3g dataset, which is based on AVHRR observations and provides a long-term global vegetation record beginning in 1982. The main objective of our study was to examine vegetation resistance and resilience across a wide range of extreme dry and wet hydroclimatic events. For this purpose, long temporal coverage was essential, because it allowed us to identify a sufficient number of extreme events across different land-cover types and SPEI categories. To our knowledge, a globally consistent EVI product directly comparable with GIMMS NDVI3g is not available for the full study period beginning in 1982. Therefore, NDVI was selected as the primary vegetation activity metric in the main analysis.

Nevertheless, we fully agree with the reviewer that the use of NDVI may influence the results and that an EVI-based sensitivity check would strengthen the interpretation. To address this point, we conducted an additional comparison using MODIS data, which provide both NDVI and EVI for the same period and spatial framework. Specifically, we recalculated resistance and resilience using MODIS NDVI and MODIS EVI and compared the resulting values across land-cover classes, SPEI timescales, and formula types. These results have been added to the Supplementary Materials as new supplementary figures (Figure S2 and S3).

Because the MODIS-based analysis covers a different time period from the GIMMS-based main analysis, and because the SPEI values used to define normal climatic conditions and event years necessarily differ between the two analyses, the MODIS NDVI–EVI comparison cannot be interpreted as a direct replication of the main results. Instead, we present it as a sensitivity analysis to evaluate whether the choice of vegetation index can affect the magnitude and direction of estimated resistance and resilience.

The MODIS comparison showed that NDVI- and EVI-based estimates were broadly similar in many land-cover classes, especially when using the formula adopted in the present study. However, some land-cover-dependent differences were also observed. In particular, for sparsely vegetated classes such as open shrublands and grasslands, resistance calculated from NDVI tended to be higher than that calculated from EVI. This suggests that, if a comparable long-term EVI product were available for the full study period, some quantitative results could differ from those obtained using NDVI alone.

We have revised the Methods and Discussion sections to clarify why NDVI was used, to acknowledge the advantages of EVI, and to explicitly discuss this limitation. We now state more clearly that NDVI was selected because of its long-term global availability, but that the use of NDVI may affect the estimated magnitude of resistance and resilience, particularly in vegetation types where saturation, soil background, or sparse vegetation effects are important. We believe that the added MODIS NDVI–EVI comparison provides a more transparent assessment of this limitation and helps readers evaluate the robustness of our conclusions.

Main manuscript: p. 8, lines 192-p. 9, lines 203: p. 21, lines 437–443 and p. 23, lines 477–480.

<u>Comment 3. The interpretation of biodiversity-related predictors is currently overstated</u>

We thank the reviewer for this important comment. We agree that the original manuscript overstated the interpretation of ASI, ASL, and ASR and did not sufficiently clarify that these variables are not direct observations of local biodiversity.

In the revised manuscript, we have made three main changes. First, we replaced terms such as “biodiversity controls” and “biodiversity drivers” with more cautious expressions, such as “biodiversity-related proxy variables,” “modeled species-change variables,” or “anthropogenic species-change proxies.” We now clearly state that ASI, ASL, and ASR represent modeled anthropogenic species increase, loss, and richness, rather than direct field observations of biodiversity.

Second, we revised the Results, Discussion, and Conclusion to avoid causal language. We no longer interpret the importance of ASI, ASL, and ASR as evidence that species gain or loss directly caused changes in resistance or resilience. Instead, we describe these results as predictive associations or predictive contributions within the LightGBM models. Any possible mechanistic interpretation is now presented as a hypothesis rather than as a demonstrated causal process.

Main manuscript: p. 20, lines 424–9. 21, lines 430; p. 22, lines 464–471 ;p. 23, lines 504.

<u>Comment 4. The thresholding of “interpretable models” is arbitrary and may bias conclusions</u>

We thank the reviewer for this important comment. We agree that the thresholds of $R^2 > 0.5$ and variation in resistance/resilience > 0.01 were not sufficiently justified in the original manuscript. We also agree that using these thresholds too strongly may have given the impression that some systems lacked interpretable controls simply because they fell below arbitrary cutoffs.

In the original analysis, these thresholds were introduced as pragmatic screening criteria to identify model combinations suitable for focused discussion among the 462 model configurations. The R^2 threshold was used to avoid over-interpreting models with weak predictive performance, whereas the variation threshold was used to avoid emphasizing cases in which the response variable showed very little variation. However, we agree that these values should not be treated as strict boundaries separating interpretable and non-interpretable systems.

To address this concern, and also in response to the reviewer's minor comment No.4 that uncertainty should be reported more systematically, we re-ran the LightGBM analyses using multiple model settings. Specifically, for each model configuration, we used three random seeds and two hyperparameter settings, resulting in six independent runs per model. This procedure was applied to all 462 model configurations, resulting in 2,772 model runs. In addition, we repeated this analysis using three learning-rate settings: 0.1, 0.2, and 0.3. In total, we therefore conducted 8,316 model runs.

Based on these repeated analyses, we summarized the robustness of the selected top predictor. We identified variables that were selected as the most important predictor across all three learning-rate settings. When the same variable was selected in only two of the three settings, we marked it with a triangle symbol. When different variables were selected in all three settings, we denoted the result as "nan." In the revised table, we also highlighted in yellow the cases in which the selected predictor differed substantially from the previous result in terms of broad variable category, such as climate, land use, biodiversity-related proxies, or soil environment.

We have also revised the interpretation of the thresholds. The thresholds are now described as heuristic criteria used for summarizing and highlighting results, rather than as rigid criteria for determining whether ecological controls exist or do not exist. These revisions make the analysis less dependent on arbitrary thresholding and provide a more transparent assessment of uncertainty across the full set of model configurations.

Main manuscript: p. 11, lines 253–p. 12, lines 294 : p. 14, lines 331–336. Table 2 note on triangle/nan: p. 19, lines 380-383.

Land Cover	SPEI	Resistance							Resilience						
		EXD	SED	MOD	NOR	MOW	SEW	EXW	EXD	SED	MOD	NOR	MOW	SEW	EXW
Evergreen Needle leaf Forest	3	T△	T	T-1△	T-1	T	T-1	T	EL	T-1	T-1	T+1	T	T+1△	T△
	6	T	T	T△	T△	T	T-1	T-1△	T-1	T-1	T-1	nan△	T+1	T+1	T
	12	T-1	T△	T-1△	T-1	T-1△	T	T-1	T-1	T-1△	T-1	T+1	T+1	T	T△
Evergreen Broadleaf Forest	3	T	T	T	T-1△	T-1	nan△	T	T-1	T-1	T-1	T-1	T-1△	T-1	T
	6	T-1	T	T△	T-1	T	P△	T-1	T-1	T-1	T-1	T-1△	T-1	T-1	T-1
	12	T-1	T-1	T-1	T	T△	T	T	T-1	T-1	T-1	T	T-1	T-1	T-1△
	3	R	T	T-1	T	R-1	R	T-1	T	T	T+1	T+1	R+1	T-1	T+1

Deciduous Needleleaf Forest	6	T	T-1	T	T-1	R-1	R	T-1	R+1△	T	T+1	T+1△	R+1	T-1	T+1
	12	T	T	T	T	T-1	R	R	P+1	T	T	R+1	T-1	T-1	T-1
Deciduous Broadleaf Forest	3	T	T	T	T	T	T	T	T△	T-1	T+1	T+1	P	P	T+1△
	6	T	T	T	T	T	T	T-1△	T-1△	T+1△	T+1	T+1	T+1△	T-1△	P△
	12	T	T	P-1△	T	T	T	T	R△	P-1	T+1△	T+1	T+1△	P	P
Mixed Forest	3	T	T-1	T-1	T-1	T-1	T-1	T△	T	T	T-1	T-1△	T+1	T+1	T+1
	6	T-1	T-1	T-1	T-1	T△	T△	T	T	T	T-1	T+1	T+1	T+1	T+1
	12	T-1	T-1	T-1△	T-1	T	T-1	T-1	T	T△	T-1	T+1△	T+1	T+1	T+1
Open Shrubland	3	ASL	T△	T-1	T-1	T	T	T	T-1	T	T+1	T	T+1	T+1	R+1
	6	nan△	P	T	T-1△	T	T	T	T+1	T	P+1	T+1	T+1	T+1	T△
	12	T	T-1	T	T-1	T-1△	T-1△	T	T-1	T	T	T+1△	T△	T+1	T-1△
Woody Savanna	3	T-1	T	T	T	T	T	ASR	T-1	T-1	T-1	T+1	T-1△	T-1	T+1
	6	T-1	T	T	T	T	T	ASR	T-1	T-1△	T+1	T+1	T+1△	T-1	T+1
	12	T-1	T-1	T	T	T	T	R-1	T-1	T△	T-1△	T+1	T-1	T+1	T+1△
Savanna	3	T-1△	ASL△	T	T	T	T	T	T-1	T-1	T+1	T+1	T+1	T-1	T
	6	T	ASL△	T	T	T	T	T	T-1	T-1	T+1△	T+1	T△	T-1	T
	12	T	ASR△	T	T	T	T	T	T-1	T-1△	T+1	T+1	T+1	T+1△	T+1△
Grassland	3	ASL	EL	T-1	T-1	T-1	T-1	EL	T-1△	T-1△	T-1	T-1△	T-1△	T-1	T
	6	EL	EL	T-1	T-1	T-1	T-1	T-1	T-1	T	T△	T	T-1	T-1△	T
	12	ASL	T-1	T-1	T-1	T-1	T-1	T△	T-1	T	T-1△	T	T-1	T-1△	T
Cropland	3	ASI	T-1	T	T	T-1	T-1	T	T+1	T-1	T-1	T	T+1	T△	T
	6	ASI	ASI	T	T	T	T	T-1	T+1	T-1	T-1	T+1△	T+1	T-1△	T

	12	ASI	T-1	T	T-1	T	T	T	T+1△	T-1	T-1	T-1△	T-1	T-1△	T△
Cropland/Natural vegetation mosaic	3	P-1	T-1△	T△	T	T-1	T-1△	T-1	T-1△	T-1	T-1	T	T-1	T-1	T+1
	6	ASR	P-1△	T	T	T	T△	T-1△	T+1	T-1	T-1	T	T+1	T+1	T+1△
	12	T-1	T-1	T	T	T	T-1	T	T-1	T-1	T-1	T-1	T+1	T+1△	T+1

5. The paper needs clearer separation between pattern, prediction, and mechanism

We thank the reviewer for this helpful comment. We agree that the original manuscript did not always clearly distinguish empirical patterns, predictive associations, and possible mechanisms. In the revised manuscript, we have removed or weakened statements that were not directly supported by the data. For example, interpretations suggesting that species loss contributes to reduced drought resistance or that the introduction of drought-tolerant crop varieties shapes ecosystem stability are now presented only as possible hypotheses, rather than as demonstrated mechanisms.

We have also revised the Discussion and Conclusion to emphasize that our study identifies broad global patterns in vegetation resistance and resilience and evaluates predictive relationships using LightGBM models. We now explicitly state that these analyses do not test the underlying ecological mechanisms. This limitation has been added to the Discussion, and speculative mechanistic explanations are clearly separated from the results directly supported by our data.

Main manuscript: p. 21, lines 433–456 ; p. 22, lines 464–471.

Response to Minor Comment 1

We thank the reviewer for pointing out this inconsistency. This was a typographical error in the original manuscript. The intended spatial resolution was 5 arc-min, not 5° × 5° . We have corrected this statement in Section 2.1 and checked the manuscript to ensure that the spatial resolution is consistently described as 5-min resolution throughout the revised version.

Main manuscript: p. 4, line 99;

Response to Minor Comment 2

We thank the reviewer for pointing out this error. This was a mistake in the notation of the original manuscript. In the normalization equation, "mean" should have been "min." We have corrected X_{mean} to X_{min} and revised the accompanying variable definitions to ensure that the min–max normalization formula is described correctly and consistently.

Main manuscript: p. 10, lines 230–244.

Response to Minor Comment 3

We thank the reviewer for pointing this out. We agree that the phrase "physiologically plausible range of 0.8–1.2" was not sufficiently clear in the original manuscript.

In the revised manuscript, we have clarified that the range of 0.8–1.2 was used only as an axis range for boxplot visualization, because our resistance and resilience indices are centered around 1. Values close to 1 indicate vegetation activity comparable to the normal-year condition, whereas values below or above this range represent relatively large deviations from the normal condition. We have revised the text to avoid implying that values outside 0.8–1.2 are physiologically impossible.

To evaluate whether this visualization range could obscure a substantial fraction of the data, we quantified the proportion of values outside the 0.8–1.2 range for each land-cover type and SPEI category. These results were summarized as sample-size-weighted percentages, calculated as the total number of values outside the range divided by the total number of valid pixels. Across all analyses, only 0.416% of values fell outside the 0.8–1.2 range. When examined separately by SPEI timescale, the corresponding percentages were 0.418% for SPEI-3, 0.417% for SPEI-6, and 0.412% for SPEI-12. Even when examined by event category, the integrated outside-range percentage was generally below 1%, with the maximum value being 1.183% for resistance under the SPEI-3 HYW condition.

We have also clarified those values outside the 0.8–1.2 range were not removed from the dataset (for LightGBM) and were retained in all statistical analyses and model fitting. The range was used only to improve the readability of the boxplots, and the frequencies of values outside this range are now reported in the Supplementary Materials (Table S3). Therefore, although a very small number of extreme values are not displayed within the boxplot axis range, this visualization choice does not affect the numerical analyses or the conclusions of the study.

Main manuscript: p. 13, lines 307–310 Supplement: Table S3, p. 15, lines 207–233.

Response to Minor Comment 4

We thank the reviewer for this comment. Because this point is closely related to Major Comment 4, we have addressed it together with that comment. In brief, we re-ran the analyses using multiple model settings and now report uncertainty in model performance and predictor importance more systematically in the revised manuscript.

Main manuscript: p. 11, lines 253–p. 12, lines 294 Model-performance summary: p. 14, lines 331–336. Table 2 note on triangle/nan: p. 19, lines 380-383.

Response to Minor Comment 5

We thank the reviewer for pointing this out. We agree that using “resilience” and “recovery” somewhat interchangeably could cause confusion. To avoid ambiguity, we have standardized the terminology throughout the revised manuscript and now consistently use “resilience” when referring to the post-event vegetation response quantified in this study.

Response to Minor Comment 6

We thank the reviewer for pointing this out. We have corrected the wording as suggested, including revising “may implicitly shape observe responses” to “may implicitly shape observed responses.” We have also carefully checked the manuscript and revised similar wording issues throughout the text.

Main manuscript: p. 2, line 51

Response to Minor Comment 7

We thank the reviewer for pointing out these wording issues in Table 1. We have corrected “Explanation of valuables in Light GBM” to “Explanation of variables in LightGBM” and revised “short radiation” to “shortwave radiation.” We have also checked Table 1 carefully and corrected similar wording and terminology issues for clarity and consistency.

Main manuscript: shortwave radiation appears correctly in Methods at p. 12, line 295 and in Table 1 .

Response to Minor Comment 8

We thank the reviewer for this helpful suggestion. In response, we created a new synthesis figure for resistance under SPEI-3, summarizing how the dominant driver classes shift with event severity across land-cover types (Figure 7). We selected this analysis because it provides a clear example of how the dominant driver classes change from near-normal conditions to extreme drought and extreme wet events.

Main manuscript: p. 22, lines 457 introduce the synthesis interpretation and refer to Figure 7;

Response to Reviewer 2

1. Conceptual framing of ecosystem stability

We thank the reviewer for this important comment. We agree that ecosystem stability is a broader concept than the resistance and resilience metrics examined in this study, and that it can also include other dimensions such as temporal variability, persistence, recovery time, and potential regime shifts.

In the revised manuscript, we have explicitly acknowledged this point as a limitation. We now clarify that our analysis captures only part of the broader ecosystem stability framework, focusing specifically on event-based resistance and one-year post-event resilience derived from NDVI dynamics. We have revised the Discussion to avoid overgeneralizing our results to ecosystem stability as a whole and to make clear that additional stability dimensions should be examined in future studies.

Main manuscript: Limitations section begins on p. 23, line 488.

2. Temporal definition of resilience

We thank the reviewer for this important comment. We agree that our definition of resilience has limitations. As the reviewer noted, recovery processes in forest ecosystems and other slow-recovering systems may extend beyond one year, and differences in ecological recovery timescales among ecosystems could introduce bias when comparing regions or land-cover types.

In this study, however, our aim was to capture as many extreme climatic events as possible using a long-term vegetation record. This consideration is also related to our choice of NDVI dataset. We used the long-term GIMMS NDVI dataset because its temporal coverage allowed us to identify a larger number of extreme dry and wet events across the global land surface. If resilience were evaluated over multiple post-event years, the corresponding number of years would be lost from the beginning or end of the study period, reducing the number of available events and limiting the comparability across event categories and land-cover types.

Therefore, we defined resilience using the NDVI in the year following the event in order to evaluate short-term post-event recovery while retaining as many event cases as possible. We have revised the Discussion to make this limitation explicit. Specifically, we now state that our resilience metric represents one-year post-event recovery and may underestimate resilience in ecosystems with slower recovery dynamics, such as forests, semi-arid systems, or ecosystems affected by severe disturbances. We also note that future studies should examine multi-year recovery trajectories to better characterize longer-term resilience.

Main manuscript: Discussion of one-year window and delayed recovery appears on p. 20, lines 420–423; limitations appear on p. 23, lines 488–491.

3. Implications of NDVI limitations

We thank the reviewer for this important comment. We agree that the limitations of NDVI, particularly saturation in high-biomass ecosystems, should be discussed more explicitly in relation to our findings.

To address this point, we conducted an additional analysis using EVI, which is designed to reduce some of the saturation effects of NDVI and to improve sensitivity in high-biomass regions. Because the GIMMS NDVI dataset used in the main analysis is based on AVHRR observations, the spectral bands required to calculate EVI are not available for the full study period. Therefore, a direct EVI-based replication of the main GIMMS-NDVI analysis was not possible.

Instead, we used MODIS data, which provide both NDVI and EVI over the same period, to compare resistance and resilience estimates derived from the two vegetation indices across land-cover types. The results of this NDVI–

EVI comparison have been added to the Supplementary Materials (figure S2 and S3).
This additional analysis showed that, in forest ecosystems, the range of values obtained using our resistance and resilience formulas tended to be narrower when NDVI was used than when EVI was used. This suggests that NDVI saturation may attenuate the apparent magnitude of vegetation responses in high-biomass ecosystems. In contrast, in sparsely vegetated systems such as open shrublands and grasslands, resistance and resilience values calculated from NDVI tended to be higher than those calculated from EVI.
We have added these points to the Discussion. Specifically, we now explain that NDVI saturation may lead to underestimation or attenuation of vegetation-response signals in dense forests, whereas differences between NDVI and EVI in sparsely vegetated systems may influence the estimated magnitude of resistance and resilience. We also note that these index-dependent differences could affect spatial patterns of inferred ecosystem vulnerability and, consequently, the relative importance of predictors in the machine-learning analysis. Therefore, the results should be interpreted with caution, particularly when comparing high-biomass forests with lower-biomass ecosystems.

Main manuscript: p. 8, lines 192- p. 9, lines 203 introduce the NDVI/EVI limitation and comparison; p. 21, lines 437–443 discuss OS and G differences; p. 23, lines 477–480 discuss canopy saturation and index-specific sensitivity.

Response to Technical Corrections

We thank the reviewer for these helpful technical suggestions.

First, thank you for your valuable suggestion. We understand that this comment arose because the section titles in the original manuscript did not clearly convey the focus of each section. While we agree that adding the suggested explanatory text could improve readability, it would also make the manuscript considerably longer. Therefore, in this revision, we have addressed the issue by revising the titles of the Results and Discussion sections so that the content and purpose of each section are more clearly indicated.

Second, we apologize for the incorrect description of the spatial resolution. The “5° × 5° ” description was a typographical error. The intended spatial resolution was 5 arc-min, not 5° × 5° . We have corrected this throughout the revised manuscript.

Main manuscript: same as Minor Comment 1: p. 4, line 99.

Third, to evaluate the robustness of the machine-learning results, we re-ran the LightGBM analyses using multiple model settings. Specifically, for each model configuration, we used three random seeds and two hyperparameter settings, resulting in six independent runs per model. This procedure was applied to all 462 model configurations, resulting in 2,772 model runs. In addition, we repeated this analysis using three learning-rate settings: 0.1, 0.2, and 0.3. In total, we therefore conducted 8,316 model runs.
Based on these repeated analyses, we summarized the robustness of the selected top predictor. We reported the variable that was most frequently selected as the most important feature across the three learning-rate settings. A triangle symbol indicates that the same variable was selected in two of the three settings, whereas “nan” indicates that different variables were selected across all three settings. These results have been added to the revised manuscript to provide a more systematic assessment of model robustness.

Main manuscript: p. 11, lines 253–p. 12, lines 294 p. 14, lines 331–336.

Finally, regarding the suggestion to mention alternative vegetation indices such as EVI, we have addressed this point in our response to Comment 3. Specifically, we conducted an additional MODIS-based NDVI–EVI comparison and added the results to the Supplementary Materials; therefore, we respectfully refer the reviewer to our response to Comment 3 for details.

Main manuscript: p. 8, lines 192-p. 9, lines 203: p. 21, lines 437–443 and p. 23, lines 477–480.