

Response to Reviewer Comments

Dear editor,

We sincerely thank you and Reviewer 2 for the careful, constructive, and highly valuable assessment of our manuscript. We greatly appreciate the time and effort devoted to evaluating the work and identifying methodological issues that required clearer explanation. The reviewer's comments were particularly helpful in strengthening the reproducibility of the media-impact workflow, the definition and interpretation of the drought-reporting sensitivity coefficient, and the distinction between drought accumulation timescale and temporal response lag.

We have considered all comments carefully and have revised the manuscript extensively. Below, we provide a point-by-point response to the reviewer's comments. For ease of assessment, each comment is reproduced, followed by our response. Where relevant, we also include the corresponding revised manuscript text in blue, together with line-number references to the revised manuscript, so that the changes can be located easily.

We are grateful for the opportunity to revise and resubmit our manuscript, and we hope that the revised version satisfactorily addresses the reviewer's concerns.

Sincerely,

Ahmed Kenawy

On behalf of all coauthors

Point-by-point response

- This study investigates the relationship between drought severity and reported agricultural and hydrological drought impacts in Spain using long-term dataset derived from corpus of El País covering the 1976 to 2023 period. LLM extracted data for impacts are consequently compared with SPI and SPEI at accumulation periods from 1 to 48 months and further investigation of drought impact sensitivity coefficient is handled using correlation and PCA analyses. Overall, I find the topic relevant for NHESS and the dataset potentially valuable. In particular, the long temporal coverage and the attempt to move beyond simple drought indicator–impact correlations towards analyzing changes in impact sensitivity across drought severity levels are interesting contributions. The manuscript also addresses an important challenge in drought research connecting hydroclimatic hazard indicators with observed societal and sectoral impacts.

We thank the reviewer for this positive overall assessment and for recognizing the value of the long-term media-derived dataset and the multiscale framework.

- However, I found several methodological steps difficult to follow and reconstruct from the current description. My main concerns relate to 3 points, specifically i) the definition and temporalspatial attribution of media-derived drought impacts, ii) the construction and interpretation of the drought-impact sensitivity coefficient, and iii) the distinction between drought accumulation timescale, temporal lag and impact response. I feel these clarifications are particularly important because the response variable represents media reporting of impacts rather than direct observations of physical impact magnitude. Some interpretations in the Results and Discussion sections appear stronger than what can currently be inferred from this data. I therefore recommend addition of further methodological clarifications and, where possible also sensitivity analysis, addressing and supporting the robustness of the main results.

Thank you for these constructive comments. In the revision, we retained the core analytical approach but sharpened the conceptual framing so that the media variable is described consistently as reporting frequency or societal visibility, rather than as a direct measure of physical damage. We made substantial methodological revisions to address the reviewer's concerns. Section 2.1 now explicitly states how article month, repeated reporting, impact categories, and provinces are assigned. Section 2.5 defines exactly which response variable enters the log–log regression, what β represents, and how temporal autocorrelation and regression residuals were assessed. We also revised the Results and Discussion to

remove causal or physical-impact interpretations that are not supported by the media-derived response variable and state explicitly that the design is contemporaneous and does not estimate an impact-response lag. The corresponding revised manuscript text is reproduced under the specific responses below.

Specific comments

- **Temporal attribution of reported impacts.** I could not clearly determine how the timing of an impact was assigned. Section 2.1 indicates that monthly impact frequencies correspond to the number of relevant articles published each month. Does this mean that the publication date was used as the date of the reported impact, or was the temporal information extracted from the article text? Please clarify how impacts were temporally attributed and discuss the implications of possible delays between impact occurrence and media reporting. If temporal information was extracted from article content, more details on this procedure and its validation would be needed.

Thanks for this comment. In the revised manuscript, we have indicated that the publication date, not temporal expressions extracted from the article body, defines the monthly position of each record. We also clarify the implication: the monthly series represent the timing of media reporting and can differ from the actual onset of a physical impact because recognition, institutional response, and publication may occur later. This is now treated as an observational limitation and is carried into the Discussion and Conclusions.

Lines 107–110: "Temporal attribution was based on the publication date of each article. Specifically, an article contributed to the calendar month in which it was published; temporal expressions within the article text were not used to back-date an impact to an inferred occurrence month. The monthly series therefore represent the timing of media reporting, which may differ from the timing of impact onset because of delays between impact occurrence, recognition, and publication."

- **Repeated reporting of ongoing impacts and independence of observations.** Related to the previous point, it is unclear how repeated descriptions of the same ongoing impacts were handled. A persistent impact such as low reservoir levels may be described in several articles across consecutive months. Were these cases treated as separate observations, or was any procedure used to identify duplicate or continuous impacts? From my understanding this could influence the further estimated drought-impact sensitivity coefficient, when long-lasting

drought may remain “newsworthy” and therefore generate reporting of the same underlying impacts.

Thanks for raising this issue. We would like to confirm that repeated reporting was retained: no event-level de-duplication was applied across articles or consecutive months. This means that persistent conditions can generate multiple media observations and therefore increase monthly reporting frequency. We now state this explicitly and interpret the response variable as reporting intensity/societal salience rather than the number of unique impact events. We also added this point to the uncertainty discussion because persistence in both drought and media attention contributes to temporal dependence.

Lines 112–115: "Repeated coverage was intentionally retained at the article level. No event-level de-duplication was applied across consecutive articles or months, so multiple reports of an ongoing condition (for example, persistently low reservoir levels) contribute multiple observations. Consequently, the resulting monthly series quantify reporting intensity or societal salience rather than the number of unique drought-impact events."

- **Spatial attribution of impacts.** In a similar manner, I would also appreciate clarification of how individual impacts were linked to individual provinces. The manuscript states that province extraction was performed separately from impact extraction. How are impact type and location subsequently connected when an article contains several impacts and/or mentions several provinces? For example, if one article describes agricultural impacts in one province and hydrological impacts in another, does the extraction procedure retain the relationship between each impact and its corresponding location, or are all extracted impact categories assigned to all provinces mentioned in the article? This could potentially introduce false province–impact combinations and is particularly relevant for the province-level analyses.

We thank the reviewer for identifying this important issue. The revised manuscript now describes the procedure exactly. Impact labels and province labels were extracted independently at the article level; the workflow did not establish sentence-level one-to-one links between a specific impact and a specific province in complex multi-impact, multi-location stories. Such articles can therefore contribute to more than one province–sector combination. We do not present these combinations as perfectly geocoded impact events; instead, the province-level analysis is now explicitly interpreted as the spatial organization of drought-related reporting. We also identify sentence-level impact–location linking as an important direction for future work.

Lines 116–121: "Province extraction was performed separately from impact extraction to identify the Spanish provinces referenced in drought-related articles. Because impact labels and province labels were extracted independently, the workflow did not establish sentence-level one-to-one links between a particular impact and a particular province when an article contained multiple impacts and multiple locations. Such articles can therefore contribute to more than one province–sector combination, and province-level results should be interpreted as spatial patterns of drought-related reporting rather than perfectly resolved geocoded impact events."

- **Definition and transformation of the impact variable used for the drought-impact sensitivity coefficient.** I found Sect. 2.5 particularly difficult to reproduce from the current description. Earlier in the Methods, monthly impact frequencies are transformed into standardized values with a mean of zero, meaning that negative impact anomalies are possible. Section 2.5 subsequently states that both drought magnitude and impact data were transformed using the natural logarithm and that observations with zero impact values were removed. It is therefore unclear to me which impact variable enters the equation. The natural logarithm cannot be applied directly to negative standardized values, and removing only zero observations would not resolve this problem. Please specify precisely whether the regression uses raw article counts, relative frequencies, standardized impact values, a shifted standardized variable, or another transformation. Any filtering or transformation applied before logarithmic transformation should be explicitly described. If observations with non-positive standardized impacts were excluded, this should also be stated and its potential influence on the estimated sensitivity coefficient discussed.

Thanks for this comment. We have rewritten Section 2.5 to remove this ambiguity. Two distinct impact representations are now explicitly separated. The standardized reporting index is used for descriptive time-series comparison, correlation analysis, and PCA. The log-log sensitivity regression does not use the standardized index; it uses the unstandardized monthly article count for the relevant province and sector. Only dry months ($\text{SPI/SPEI} < 0$) are retained; the drought index is sign-reversed to obtain a positive drought magnitude, and only positive article-count months enter the logarithmic regression because $\ln(0)$ is undefined. We also state that exclusion of zero-report months means β is conditional on at least one report being present and may be affected in sparsely reported province–sector combinations.

Lines 183–192: "For the sensitivity analysis, only months with dry index values (SPI or $SPEI < 0$) were retained, and the drought index was sign-reversed so that larger positive values represent greater drought magnitude ($Dt = -SPIt$ or $-SPEIt$). This means that the magnitude of drought was represented by the absolute value of the drought index (López-Moreno et al., 2013; Barker et al., 2016; Conradt et al., 2023). Unlike the standardized impact series used for the temporal comparison, correlation analysis, and PCA, the response variable in the log–log regression was the unstandardized monthly article count for the corresponding province and sector. Thus, I_t denotes the number of media reports in month t . Because $\ln(0)$ is undefined, months with $I_t=0$ were excluded from this specific regression step; months with positive counts were transformed as $\ln(I_t)$, and drought magnitude was transformed as $\ln(Dt)$. The resulting coefficient therefore describes the elasticity of reporting intensity conditional on at least one report being present. It does not model the probability of whether any report occurs, and exclusion of zero-report months can affect the estimated coefficient, particularly in sparsely reported province–sector combinations."

- **Robustness and uncertainty of the drought-impact sensitivity coefficient.** The proposed sensitivity coefficient is an interesting part of the study, but I have concerns about estimating it independently within relatively narrow drought-severity classes. Within categories such as 1.28–1.65, the explanatory variable spans only a limited range, while severe and extreme drought observations are presumably relatively rare. This may result in unstable slope estimates, particularly when models are further separated by province, sector, drought indicator, and accumulation period. I suggest reporting the number of observations underlying these regressions and providing uncertainty estimates (e.g. confidence intervals) for the sensitivity coefficients. It would also be useful to demonstrate that the main patterns are not driven by a small number of observations in the severe and extreme classes.

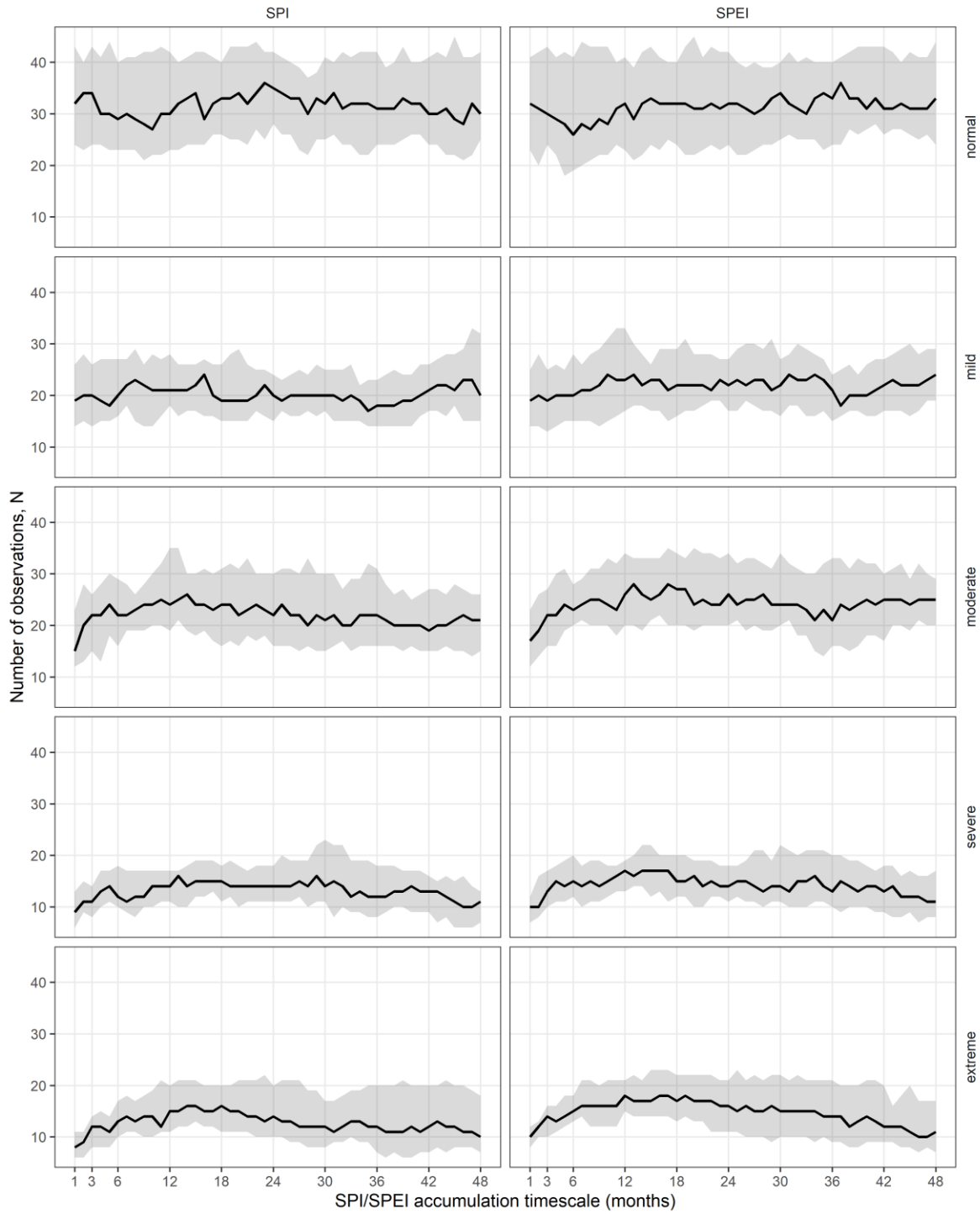
We thank the reviewer for highlighting the potential instability of sensitivity coefficients estimated within relatively narrow drought-severity classes. We agree that both the restricted predictor range and the reduced number of observations, particularly for severe and extreme droughts, can affect the stability of class-specific regression slopes. We therefore quantified the number of paired observations underlying the regressions for every drought-severity class, drought indicator, accumulation timescale, and sector and added new supplementary diagnostics (Figs. S1–S3). Figures S1 and S2 show the median number of observations and interquartile range across provinces over the complete 1–48-month accumulation spectrum for hydrological and agricultural impacts, respectively. As expected, sample sizes decline in the severe and extreme classes relative to normal, mild, and moderate drought conditions. To address the reviewer’s concern more directly, Fig. S3 focuses exclusively on severe and extreme droughts and includes reference thresholds of $N=10$ and $N=15$. The results show that median

sample sizes generally remain at or above approximately 10 observations across most accumulation timescales and are commonly around or above 15 observations over intermediate timescales. Sample support becomes weaker for some combinations at the longest accumulation periods, particularly under extreme drought, and the interquartile ranges demonstrate substantial interprovincial variability. We have therefore revised the text to make clear that individual severe- and extreme-class coefficients are interpreted cautiously. At the same time, the relatively smooth evolution of sample size across accumulation timescales indicates that the principal medium- to long-timescale sensitivity patterns are not associated with isolated timescales having exceptionally small sample sizes. Accordingly, our interpretation emphasizes coherent behaviour across neighbouring accumulation periods, indicators, sectors, and spatial units rather than individual coefficients.

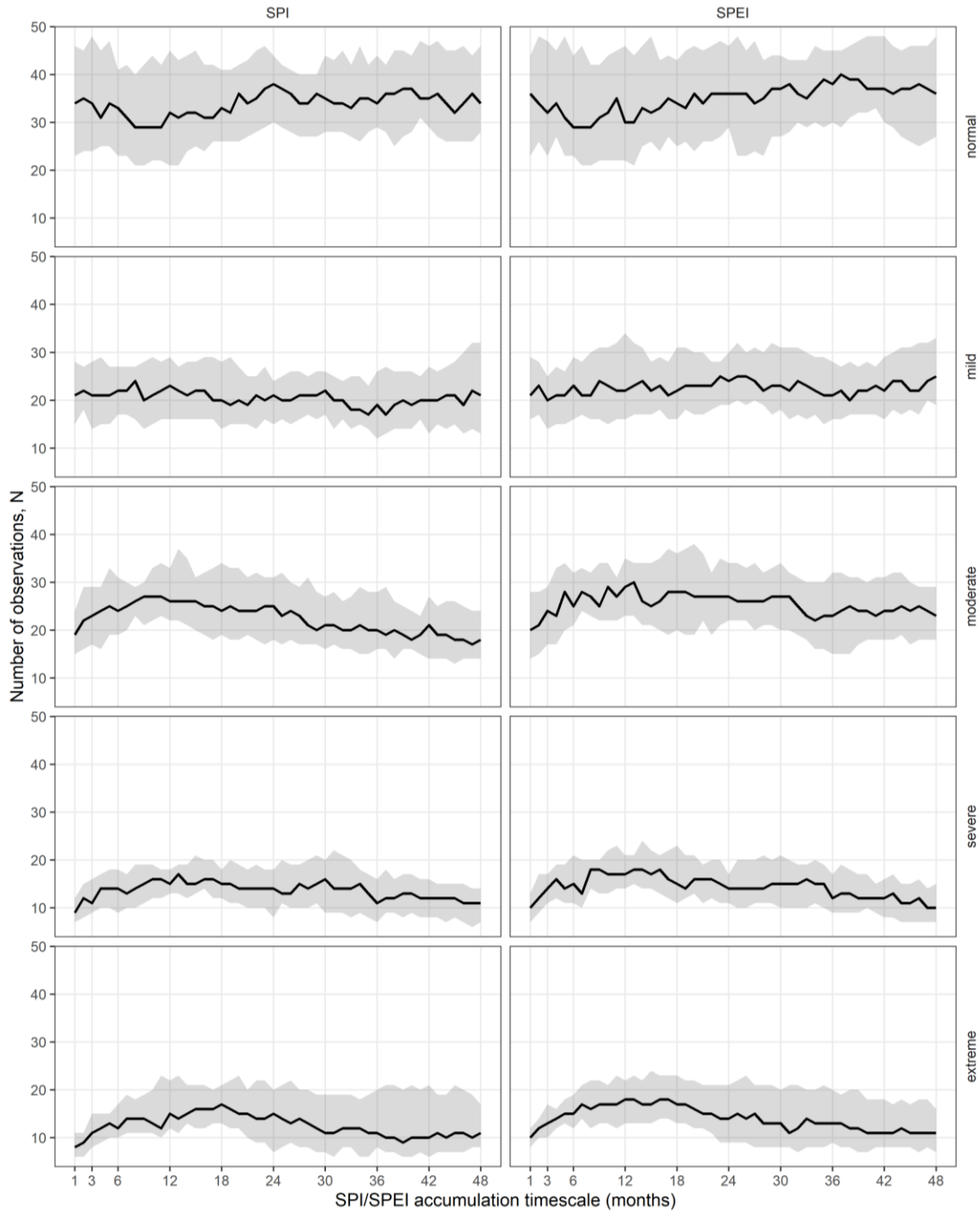
Lines 194–201: "To examine whether the reporting–drought association changes across drought severity, monthly observations were stratified by the sign-reversed drought index into near-normal dry conditions (>0 to 0.5), mild (0.5–0.84), moderate (0.84–1.28), severe (1.28–1.65), and extreme (>1.65) drought, together with a category containing all dry conditions (>0) (Agnew, 2000). Regressions were fitted separately within each class. This stratification is intended to reveal broad changes in reporting sensitivity across severity ranges, but the width of the predictor range and the number of observations differ among classes. In particular, severe and extreme classes can contain relatively few observations, so isolated class-specific slopes may be unstable. As such, the number of observations supporting each regression, as well as the resulting sample-size distributions, was quantified across provinces, accumulation timescales, drought indicators, sectors, and severity classes."

Lines 285–293: "Because stratification by drought-severity class reduces the number of observations available for individual regressions, we examined the sample size underlying the sensitivity estimates across sectors, drought indicators, and accumulation timescales (Supplementary Figs. S1–S3). As expected, for hydrological (Supplementary Fig. S1) and agricultural (Supplementary Fig. S2) impacts, the number of observations decreases from normal and moderate conditions toward severe and extreme drought classes. Nevertheless, the severe and extreme regressions retain observations across the full 1–48-month accumulation range for both SPI and SPEI and for both hydrological and agricultural impacts. Median sample sizes generally remain around or above 10 observations across most timescales, and frequently approach or exceed 15 observations over the intermediate accumulation periods where many of the strongest sensitivity responses occur (Supplementary Fig. S3)."

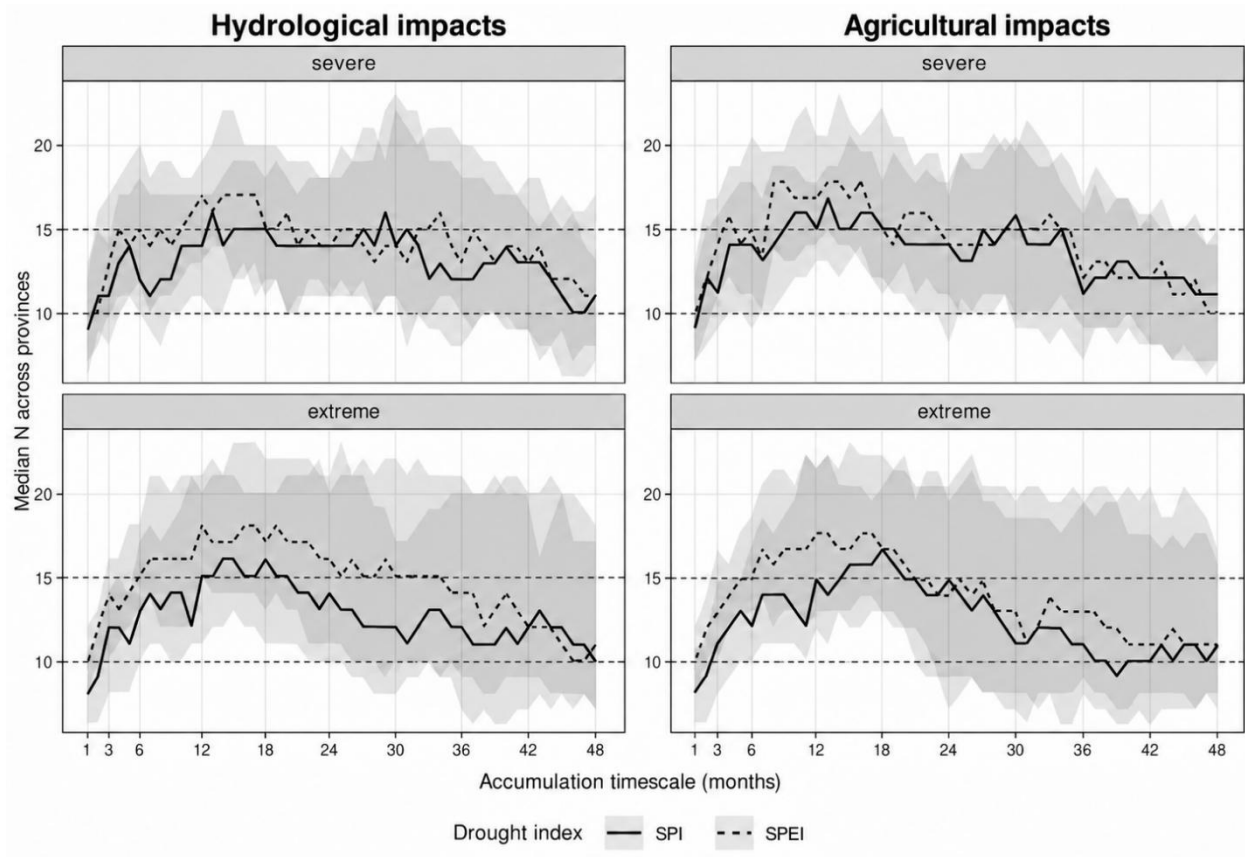
Lines 295–299: "The interquartile ranges further indicate appreciable spatial variability in the number of observations available among provinces, confirming that some province–timescale regressions are less well supported than others. Importantly, however, the sample-size curves vary smoothly with accumulation timescale rather than showing isolated collapses at the accumulation periods associated with the principal sensitivity maxima. This suggests that the broad timescale-dependent patterns are not simply artefacts of a single or very small number of sparsely sampled regressions."



Supplementary Figure 1. Regression sample size across SPI (left panels) and SPEI (right panels) accumulation timescales (1–48 months) for five drought-severity classes corresponding to hydrological impacts. The black line shows the median number of paired observations (N) across 48 provinces, and the gray ribbon indicates the interquartile range (25th–75th percentiles).



Supplementary Figure 2. Regression sample size across SPI (left panels) and SPEI (right panels) accumulation timescales (1–48 months) for five drought-severity classes corresponding to agricultural impacts. The black line shows the median number of paired observations (N) across 48 provinces, and the gray ribbon indicates the interquartile range (25th–75th percentiles).



Supplementary Figure 2. Sample-size support for severe and extreme drought regressions across accumulation timescales (1–48 months). Lines show the median number of observations (N) across 48 provinces for SPI and SPEI, with gray shading indicating the interquartile range. Horizontal dashed lines mark reference sample sizes of N=10 and N=15.

- **Temporal autocorrelation and statistical independence.** Both drought indices and impact-reporting time series are likely to exhibit substantial temporal autocorrelation. This is particularly important for SPI/SPEI calculated at long accumulation periods, since successive monthly values at, for example, 24- or 48-month accumulation periods contain strongly overlapping climatic information. It is unclear whether temporal autocorrelation was considered when evaluating correlation significance or uncertainty in the regression models. Please clarify whether this issue was accounted for, and additional consideration for statistical significance and uncertainty should be also discussed.

We thank the reviewer for raising this important point. Temporal autocorrelation was examined before inferential interpretation of the correlation and regression results. Autocorrelation diagnostics were applied to the monthly drought-index and impact-reporting series, with particular attention to longer

SPI/SPEI accumulation periods because successive monthly values increasingly share overlapping climatic information. Serial dependence was assessed using the autocorrelation function (ACF), and regression residuals were additionally examined for autocorrelation to assess whether temporal dependence materially affected inference from the fitted models. The manuscript has been revised to state this procedure explicitly.

Lines 215–219: "Prior to inferential interpretation of the correlation and regression analyses, temporal autocorrelation was evaluated for the drought-index and impact-reporting series, with particular attention to the longer SPI/SPEI accumulation periods for which successive monthly values contain strongly overlapping climatic information. Serial dependence was assessed using the autocorrelation function (ACF). Autocorrelation of the regression residuals was additionally examined to assess whether temporal dependence materially affected the statistical inference from the fitted models."

- **Interpretation of media-derived impact frequency.** More generally, I think the manuscript should maintain a clearer distinction between reported impact frequency and physical impact magnitude. Article counts provide valuable information about the societal visibility, timing, and spatial distribution of drought impacts, but one article describing severe losses does not necessarily represent a smaller impact than several articles describing the same moderate impact. This distinction becomes particularly important when interpreting the sensitivity coefficient. Statements that impacts increase disproportionately with drought magnitude could be interpreted as referring to physical impact severity, whereas the analysis directly demonstrates a disproportionate increase in media-reported impact frequency. I recommend reviewing the terminology throughout the manuscript and ensuring that conclusions remain tied to what is directly represented by the response variable. This does not diminish the usefulness of the dataset, but it provides a more precise interpretation of the results.

We agree and have applied this distinction systematically throughout the manuscript. We now use terms such as “media-reported impact frequency,” “reporting count,” “reporting intensity,” and “societal visibility” when referring to the response variable. The standardized series are not described as physical impact magnitude, and β is defined as a drought–reporting elasticity rather than a physical-impact sensitivity. Statements about disproportionate increases have been revised to refer to reporting frequency, and the Discussion explicitly states that the severe/extreme patterns should not be interpreted as evidence of a physical damage threshold or resilience limit. The Conclusions now maintain the same distinction.

Minor and technical comments

- Please check terminology related to “drought duration”, “accumulation period”, “response time”, and “lag”, as these currently appear to be used somewhat interchangeably in parts of the manuscript.

We agree. We reviewed the manuscript systematically and now use “accumulation period/timescale” only for the window over which SPI/SPEI is calculated, “drought-event duration” for the persistence of a drought event, and “response/reporting lag” for a temporal offset between climatic conditions and a later impact/report. Because the present analysis does not estimate the latter, we explicitly state this throughout.

Lines 227–231: "In addition, the dominant accumulation timescale was defined as the timescale for which the sensitivity coefficient was the maximum in the range 1–48 months for each province and drought severity class. All drought indices were aligned with media reporting in the same month; therefore, the identified accumulation timescale represents the climatic integration window most strongly associated with reporting at time (t), rather than drought-event duration or an explicit impact-response lag."

Lines 432–438: "Importantly, although newspaper reports often describe drought consequences when they become observable or socially salient, the temporal delay between physical impact occurrence and publication was not quantified in this study. For many reported impacts, this delay may be relatively short compared with the multi-month climatic accumulation windows examined here; however, drought impacts are often gradual and persistent, and some reports may describe conditions that began earlier. We therefore use publication month as the timing of reporting rather than as an exact estimate of impact onset. Accordingly, any temporal offset between climatic conditions and the media record may reflect a combination of physical drought-impact response time, recognition of the impact, and reporting delay."

- Please state clearly whether one article can contribute to both agricultural and hydrological categories.

We have now stated this explicitly. The extraction is multi-label rather than mutually exclusive, so an article that reports both agricultural and hydrological consequences contributes to both broad sectoral series. We also retain this caveat when interpreting the strong correlation between the two series.

Lines 128–131: "The extraction was multi-label rather than mutually exclusive. Therefore, a single article could contribute to more than one impact category when it reported multiple drought consequences, such as crop losses together with reservoir depletion or water-use restrictions. The agricultural and hydrological series should therefore be interpreted as sector-specific monthly reporting frequencies, not as disjoint sets of articles."

- The aggregation of crop and livestock impacts into “agricultural impacts” and water-resource and energy impacts into “hydrological impacts” should be further justified. These components may have substantially different response mechanisms and timescales.

Thanks for this comment. The revised Methods now explains that these are operational broad-sector groupings introduced to reduce sparsity and allow consistent province-by-timescale comparison; they do not imply identical mechanisms or response timescales among the constituent categories. We explicitly note that crop and livestock impacts can differ from one another, as can water-resource and energy impacts, and therefore the sector-level results should be interpreted as broad reporting signals.

Lines 103–106: "These aggregations were introduced to reduce data sparsity and support consistent province-by-timescale comparisons at a broad sectoral level. They represent operational groupings rather than an assumption that all constituent impacts share identical mechanisms or response timescales. The resulting sector-level patterns should therefore be interpreted as broad reporting signals."

- Please clarify what is meant by relative frequency and whether impact counts were normalized by the overall number of articles published by El País within the corresponding period. (The study covers 1976–2023, during which the volume, format, and editorial practices of newspaper publishing have changed considerably, particularly with the transition towards digital publishing. I am not fully convinced that standardizing provincial impact-frequency distributions alone removes temporal changes in reporting intensity.)

The manuscript now defines the basic variable as the monthly number of drought-related El País articles assigned to each province and sector, which we call monthly reporting frequency/reporting count. These counts were not normalized by the total number of all El País articles published in the corresponding month or year. Importantly, we no longer claim that standardization removes long-term changes in reporting intensity. Instead, we state explicitly that standardization improves comparability of the distributions but cannot remove changes in newspaper production, editorial priorities, digital publication practices, public attention, or drought newsworthiness. This limitation is also emphasized in the Discussion and Conclusions.

Lines 123–125: "The final dataset is a monthly count of articles referring to each impact category (agriculture and hydrology) in each province. Monthly impact frequencies were quantified as the number of drought-related articles published for each province and impact category."

Lines 161–163: "However, standardization does not remove possible long-term changes in newspaper production, editorial priorities, digital publication practices, or drought newsworthiness. These sources of non-climatic temporal variation are therefore treated as limitations of the media-derived record rather than assumed to have been eliminated by standardization."

- The spatial heterogeneity shown in the results is interesting, but I think the Discussion could go further in considering why provinces differ in their drought-impact relationships.

We agree and substantially expanded this part of the Discussion. The revised text now considers climatic gradients and large-scale circulation, basin storage and groundwater exchange, irrigation dependence, crop composition, water-allocation rules, adaptive capacity, exposure, and the geography/intensity of newspaper coverage as plausible contributors to provincial differences. At the same time, we explicitly avoid causal attribution because these variables were not included as predictors in the statistical models. The spatial patterns are therefore interpreted as empirical heterogeneity in the drought–reporting relationship rather than causal maps of physical vulnerability.

Lines 472–482: "Several processes could plausibly contribute to the observed geography. Large-scale atmospheric circulation structures drought variability across Spain. Previous studies have linked drought variability in Spain to large-scale atmospheric circulation patterns (Vicente-Serrano et al., 2006; Russo et al., 2015; Bakke et al., 2023; Noguera et al., 2024). However, basin storage, groundwater exchange, irrigation dependence, crop composition, water-allocation rules, and adaptive capacity can alter how the same climatic deficit is experienced and reported. Numerous studies demonstrate that local hydro-climatic conditions, water management practices and socio-economic factors modulate drought responses, with regional differences in sensitivity and dominant timescales (Grau-Satorras et al., 2016; Urquijo et al., 2015; Estrela et al., 2016; Iglesias et al., 2021; Junquera et al., 2025). In addition, province-specific reporting intensity and the geographic focus of a national newspaper may contribute to apparent spatial contrasts. The spatial heterogeneity therefore likely combines hydro-climatic, management, exposure, and reporting effects and should not be attributed to climate alone."

Lines 524–526: "These factors provide plausible physical and management contexts for the regional differences observed here, but the media-based analysis cannot separate their individual contributions from differences in exposure or reporting. Accordingly, spatial variation in the sensitivity coefficient (β) can be interpreted as variation in the drought–reporting relationship rather than as a direct map of physical vulnerability."

Finally, we again thank the editor and Reviewer 2 for the careful and constructive comments. The review prompted substantial improvements in the methodological transparency, statistical interpretation, terminology, and scope of inference of the manuscript, and we believe the revised version is considerably clearer and more robust as a result.