RC2: 'Comment on egusphere-2026-303', Anonymous Referee #2, 21 May 2026

Causal drivers of alpine flood variability from 1300 to 2020 revealed by climate time series analysis

Juan Carlos Peña-Rabadán, Lothar Schulte, Josep Ramon Miró-Cubells, Enric Casellas-Masana, and Filipe Carvalho

The manuscript addresses a highly relevant topic and contains potentially significant and interesting findings, particularly regarding the possible role of total solar irradiance in shaping flood variability. The combination of paleoflood reconstructions with causal inference methods is potentially promising and could provide useful insights into long-term climate–flood interactions. However, in its current form, the manuscript is difficult to follow methodologically and conceptually, and several aspects of the causal interpretation require more careful justification before the work can be considered robust.

Reviewer comment:

The presentation and organization of the manuscript would benefit from substantial revision. The Introduction relies heavily on cross-referencing to previous publications by the same research group and to supplementary figures and tables very early in the text, which hinders readability and makes the rationale of the study difficult to follow. The motivation for specifically examining these drivers, and particularly for moving from correlation-based analyses toward causal inference, is not sufficiently explained. Several figures are overloaded with information and contain extensive embedded text (e.g., the inclusion of research-group information in Figure 2). Also, the manuscript relies excessively on supplementary material for understanding the main analysis, while the extensive use of acronyms further reduces accessibility (e.g., AL and AH in Figure 3 are not explained in the main manuscript but only in the supplement). I encourage the authors to substantially revise the structure and presentation of the paper to improve transparency and readability.

Response:

We thank the reviewer for this constructive and detailed critique. We fully agree that the original manuscript suffered from poor readability due to excessive self-citation, early references to supplementary material, overloaded figures, and unexplained acronyms. In response, we have undertaken a **substantial revision of the entire manuscript** with a focus on transparency and accessibility. Below we address each specific point.

- ✓ **Introduction: reduced self-citation and improved motivation for causal inference**

  We have completely rewritten the Introduction in four subsections:

  1.1 The challenge of flood attribution in a changing climate
  1.2 The limitations of correlation-based approaches in paleoflood hydrology
  1.3 Specific challenges of paleoclimate proxies for causal inference
  1.4 The emergence of causal inference in Earth system science

- **Provide a clear, step-by-step justification** for moving from correlation to causal inference, including:
    - Three specific limitations of correlation-based attribution (confounding, non-stationarity, directionality). *Lines 29-57 in section 1.2* in the revised manuscript.
    - A dedicated subsection on the specific challenges of paleoclimate proxies (temporal resolution, dating uncertainties, smoothing, autocorrelation, proxy sensitivity changes). *Lines 58-78 in section 1.3* in the revised manuscript.
    - A review of recent causal inference applications in Earth system science (Runge et al., 2019; Kretschmer et al., 2016; Su et al., 2023; Silini et al., 2023, etc). *Lines 79-111 in section 1.4* in the revised manuscript
    - An explicit statement of the gap: no previous constraint-based causal discovery applied to a multi-century Alpine paleoflood record. *Lines 92-99 in section 1.4* in the revised manuscript: *"Despite these advances, the application of formal causal inference to long-term (centennial to millennial) flood records remains remarkably rare. To our knowledge, no study has systematically employed a constraint-based causal discovery algorithm on a multi-century paleoflood archive from the European Alps to identify the drivers of summer flood variability. A few recent studies have applied Granger causality to flood-climate relationships using instrumental or historical records (e.g., Šraj et al., 2016; Mediero et al., 2014), but these typically cover only the last 50–150 years and do not account for the specific challenges of paleo-data (dating uncertainty, variable resolution, proxy non-stationarity). This gap is surprising given the rich paleoflood and paleoclimate datasets available for the Alps (Wirth et al., 2013; Glur et al., 2013; Schulte et al., 2015) and the urgent need for better attribution of natural vs. anthropogenically driven changes in flood risk (Wilhelm et al., 2022; Blöschl et al., 2020)."*
- **Minimise cross-references to previous work by our group** while retaining essential references to the paleoflood reconstruction (Schulte et al., 2015, 2019) and the SNAO definition (Peña et al., 2015) as data sources, not as the sole intellectual framework.
  New *section 2 Study area and previous advances: the Hasli-Aare catchment paleoflood record* with five subsections:

  2.1 Geographical and climatic setting
  2.2 The AA-05 sedimentary record: a multi proxy paleoflood archive
  2.3 Previous reconstructions and their empirical foundation
  2.4 Limitations of previous correlation-based interpretations
  2.5 From correlation to causation: the present contribution

- **Remove early references to supplementary figures and tables**; all supplementary material is now cited only after the reader has been introduced to the concepts.

✓ **Figures: reduced clutter and improved clarity**

- Figure 3 (now Figure 4 – the RPCA teleconnection patterns): All acronyms (AL, AH, OM, EATL/WRUS, EA) are now defined in the figure caption and in the main text (Table 2, glossary).
- We have also added a sentence in Section 3 (Climate Variables) explicitly listing these patterns and referring the reader to Table 2.

✓ **Reduced reliance on Supplementary Material**

We have moved **key descriptive content** from the *Supplementary Material* into the main text:

- **Section 2** now includes a detailed description of the AA-05 sedimentary record including a summary table (Table 1) with core characteristics, dating methods, temporal resolution, and proxy limitations.
- **Section 3** now includes the glossary of all variables (formerly Table S1) as Table 2 in the main text.
- The description of SLP grids (formerly Table S2) is now also in the main text (Table 3).

New section was created: *3 Climate Variables*, where a first point that described preprocessing and harmonisation of the variable, and after we described the variables used in the analysis divided Tiers based on the posterior causal inference.

3.1 Data preprocessing and harmonisation
3.2 External forcings (Tier 1)
3.3 Large-scale atmospheric variability (Tier 2)
3.4 Regional atmospheric modes (Tier 3)
3.5 Regional climate responses (Tier 4)
3.6 Hydrological impact (Tier 5)

Consequently, the main text is now self-contained for all essential information; the *Supplementary Material* is reserved for extended methodological details (e.g., full Granger causality tables, PC-algorithm technical notes, bootstrap sensitivity analyses) and additional figures.

✓ **Acronyms: full definitions in the main text**

All acronyms used in the main text and in figures are now defined at first use and compiled in Table 2 (glossary). Specifically:

- **AL (Atlantic Low)**, **AH (Atlantic High)**, **OM (Omega)**, **EATL/WRUS (East Atlantic/Western Russia)** and **EA (East Atlantic)** appear in Table 2 (formerly Table S1.1) with their full names and sources.
- We have checked that no acronym appears in the main text without an accompanying definition.

Reviewer comment:

The methodological assumptions underlying the causal framework also require further scrutiny and discussion. In particular, the hierarchy imposed and expert-defined blacklist constraints used to guide the causal graph construction raise concerns regarding circularity. While some excluded causal directions may indeed be physically implausible (e.g., floods affecting solar forcing), the extent to which the inferred structure depends on these imposed assumptions remains unclear. Given the complexity and multiscale coupling of the climate system, the assumptions embedded in the blacklist design require more explicit justification. The imposed blacklist constraints and their influence on the resulting network topology and causal interpretation should be reported and discussed more explicitly.

Response:

We thank the reviewer for this important methodological critique. We agree that the use of expert-defined blacklists and a tiered hierarchy requires transparent justification and a quantitative assessment of how much the inferred structure depends on these constraints. In the revised manuscript, we have:

- ✓ **Explicitly justified each blacklist constraint** with references to physical principles and the temporal scales of our study (centennial to multi-decadal).
- ✓ **Quantified the impact** of the blacklist by comparing the unconstrained and constrained DAGs (number of edges removed, edges re-oriented, core pathways preserved).
- ✓ **Added a sensitivity analysis** using a less restrictive blacklist to show that the main causal pathway (TSI → EOF1 → Pamj_Alps → Flood_F1) is robust to the specific choice of constraints.
- ✓ **Discussed the potential for circularity** and clarified that our constraints are based on *established domain knowledge* (e.g., external forcings cannot be caused by internal variables on these timescales), not on the data itself.

Below we provide the specific changes made to the manuscript and a summary for the reviewer.

Changes made in the revised manuscript

1. In *Section 4.5 Incorporating Domain Constraints in Causal Inference*: expanded and justified

We have replaced the original short paragraph with a more detailed justification:

Lines 342-351 in section 4.5

"The blacklist includes the following categories of prohibited edges: (i) any edge from a lower-tier variable (e.g., flood index, Alpine temperature) to a higher-tier variable (e.g., TSI, NH_Volcanic), because retro causality is physically impossible; (ii) edges from regional responses (e.g., Pamj_Alps, Tjja_Alps) to large-scale circulation indices (e.g., EOF1, SNAO), because local precipitation/temperature cannot drive hemispheric circulation on the timescales considered (multi-decadal to centennial); (iii) edges among external forcings (TSI ↔ NH_Volcanic), as solar and volcanic forcing are independent on

these timescales (Lean, 2022; Sigl et al., 2015). These constraints are conservative; they only exclude directions that would contradict fundamental physical laws or well-established observational evidence (Spirtes et al., 2000; Peters et al., 2017). The tiered architecture and use of blacklists together serve as a form of soft expert knowledge integration, which has been shown to improve the robustness and realism of causal inference in Earth system sciences (Runge et al., 2019; Kretschmer et al., 2021).”

2. In *Section 5.4 Key findings from the constrained DAG*: we added a sentence quantifying the impact

We have replaced the original short paragraph with a more detailed justification:

Lines 425-441 in section 5.4

“The influence were assessed of the blacklist on the final DAG. We compared the unconstrained and constrained networks. The blacklist removed 23 % of the edges present in the unconstrained skeleton – exclusively those that violated the tiered ordering (e.g., Pamj_Alps → TSI). It did not remove any edge that was part of the core pathway TSI → EOF1 → Pamj_Alps → Flood_F1. The constrained DAG retained all edges that survived the bootstrap stability threshold and that were consistent with the hierarchy. As a further sensitivity test, we ran the PC stable algorithm with a reduced blacklist that allowed, for example, temperature affecting volcanic activity (which is physically unrealistic). This less restrictive blacklist produced a DAG with 14 % more edges, but the core pathway remained unchanged. This indicates that the main causal findings are robust to the specific blacklist design, provided that the most fundamental physical constraints (e.g., no retro causality to external forcings) are respected. The results of this sensitivity analysis are reported in Supplementary Material S2.

The constrained DAG resulted in a lower significance threshold for edge retention (0.41; Fig. S1. and Table S2, in Supplementary Material S2), while preserving key structural components of the unconstrained network. A visual comparison of the unconstrained (Fig. 6) and constrained DAGs (left panel in the Fig. 7) highlights how the integration of physical knowledge alters the inferred network topology and strengthens interpretability. Although the constrained DAG uses a lower inclusion threshold (0.41), sensitivity analysis (Supplementary Material S2) demonstrates that the core causal relationships are also recovered with stricter thresholds (0.7, Table S4); weaker edges that drop out (e.g., TSI→flood_time) are not central to our attribution. No edge belonging to the core pathway TSI → EOF1 → Pamj_Alps → Flood_F1 was removed, reinforcing the links detected by the Granger Causality Test.”

3. New *Section 4.7 Treatment of autocorrelation and low-frequency variability*

Lines 366-381 in section 4.7

“Paleoclimate time series often exhibit strong autocorrelation and long-term persistence, which can inflate conditional independence tests and lead to spurious edge detection (Mudelsee, 2019; Thejll & Schmith, 2005). Our causal inference framework addresses this issue through three complementary strategies. First, the bootstrap stability analysis (100 iterations with ±20-year temporal perturbations) implicitly breaks spurious temporal dependencies by resampling blocks of years, a procedure known to mitigate the effects of autocorrelation in causal structure learning (Runge et al., 2019). Second, the 21-year Gaussian

smoothing applied to key variables reduces high-frequency noise and removes short-term persistence, focusing the analysis on multi-decadal to centennial timescales where causal relationships are more robust. Third, the tiered blacklist prevents the algorithm from learning edges that could be artefacts of shared low-frequency trends (e.g., local precipitation driving hemispheric circulation).
As a sensitivity test, we attempted to pre-whiten the series using AR(1) models and applied Granger causality tests. However, diagnostic checks (Ljung-Box test) revealed that AR(1) was insufficient to remove autocorrelation in most variables ($p < 0.05$). Therefore, we do not rely on pre-whitened Granger results as primary evidence. Instead, we performed additional sensitivity tests using partial correlations (reported in Section 5.5 and Supplementary Material S2.4) and higher-order ARIMA models selected by the Akaike Information Criterion. The core causal pathway remained stable under those more complex pre-whitening schemes, confirming that our conclusions are not artefacts of autocorrelation."

4. New section 5.5 Sensitivity to autocorrelation and low-frequency variability

Lines 463-480 in section 5.5

"To assess whether autocorrelation or long-term persistence could affect our conditional independence tests, we performed a partial correlation analysis (Supplementary Material S2.4, Table S5). For each pair of key variables (TSI, EOF1, Pamj_Alps, Flood_F1, NH_Volcanic, Tjja_Alps), we computed the partial correlation while conditioning on all other variables. The results (Table S5) show:

- *The link TSI → EOF1 is not significant in the partial correlation test ($r = 0.026$, $p = 0.497$). This does not contradict the causal link recovered by the PC-stable algorithm, because the latter operates on smoothed (21-year) data and incorporates domain constraints that reveal the multi-decadal influence of solar forcing on atmospheric circulation (Gray et al., 2010; Ineson et al., 2011).*
- *Conversely, the significant negative partial correlation between TSI and Flood_F1 (-0.460, $p < 0.001$) supports the direct causal pathway from solar activity to flood variability identified in the main analysis.*
- *The core causal pathway (TSI → EOF1 → Pamj_Alps → Flood_F1) remained stable under those more complex pre-whitening schemes, confirming that our conclusions are not artefacts of autocorrelation.*

Pre-whitened Granger causality tests using AR(1) models were also explored, but diagnostic checks (Ljung-Box test) showed that AR(1) was insufficient to remove autocorrelation in most variables ($p < 0.05$). Therefore, those results are not reported here. The core causal pathway remains robustly supported by the bootstrap-stabilised PC-stable algorithm with 21-year smoothing, which is less sensitive to residual autocorrelation (see Section 4.7). We conclude that our main causal findings are not artefacts of autocorrelation or long-term persistence."

5. In the *Supplementary Material 2*: *new sensitivity analysis (S2.3) and Robustness to autocorrelation and low-frequency variability (S2.4)*

We have added two subsections (as also requested in the response to Reviewer 1 that includes:

- In the *S2.3. Sensitivity to bootstrap threshold*. New Table S4. Edge stabilities (bootstrap frequencies) for key directed edges at different thresholds. And added the paragraph *(lines 173-180 in Supplementary Material 2)*:

  "To test whether our main conclusions depend on these specific thresholds, we re-ran the causal discovery using stricter thresholds of 0.7 and 0.8 for both DAGs. Table S4 reports the edge stabilities (bootstrap frequencies) for the core pathways at these higher thresholds.
  The central causal pathway TSI → EOF1 → Pamj_Alps → Flood_F1 remains intact at all thresholds (stabilities >0.7 for TSI→EOF1, >0.75 for EOF1→Pamj_Alps, >0.7 for Pamj_Alps→Flood_F1). Weaker edges, such as TSI→flood_time (0.57) and EOF1→flood_time (0.44), drop out at the stricter thresholds, but these are not essential for our main interpretation. Thus, the overall causal narrative is robust to the choice of threshold."

- New *Section S2.4 Robustness to autocorrelation and low-frequency variability*. We added the paragraph *(lines 183-192 in Supplementary Material 2)*:

  "This section provides the partial correlation analysis discussed in Section 5.6 of the main text. Pre-whitened Granger causality tests using AR(1) models were also explored, but diagnostic checks (Ljung-Box test) showed that AR(1) was insufficient to remove autocorrelation in most variables ($p < 0.05$). Therefore, those results are not reported here. The core causal pathway is robustly supported by the bootstrap-stabilised PC-stable algorithm with 21-year smoothing, which is less sensitive to residual autocorrelation. Partial correlations. For each pair of key variables (TSI, EOF1, Pamj_Alps, Flood_F1, NH_Volcanic, Tjja_Alps), we computed the partial correlation while conditioning on the remaining four variables, using the `pcor.test` function from the ppcor R package. Table S5 reports the partial correlation coefficients and their associated p-values."

Reviewer comment:

Related to this point, the manuscript does not sufficiently discuss how autocorrelation, long-term persistence, and shared low-frequency variability among climate series may affect conditional independence testing and inflate apparent causal structure. This is particularly important in paleoclimate applications where proxy smoothing, and persistence may induce shared covariance patterns even in the absence of direct mechanistic causality.

The interpretation of structural differences between the pre-industrial and industrial periods also appears somewhat overstated. The manuscript repeatedly interprets changes in DAG topology as evidence of anthropogenic restructuring of the climate system. However, no explicit anthropogenic variables are included in the network, and the term "anthropogenic effects" remains rather vague throughout the text. It is unclear whether the authors refer to recent warming, land-use changes, river engineering, or broader climatic shifts. Alternative explanations such as differences in data quality, proxy uncertainty, sample size, or methodological instability are not sufficiently examined. Therefore, more caution is

required when interpreting edge appearance/disappearance as evidence of anthropogenic reorganization of climate–flood relationships.

Response:

We thank the reviewer for raising these two important and interconnected issues. We fully agree that autocorrelation and persistence can inflate conditional dependence tests, and that our previous interpretation of structural shifts as "anthropogenic" was insufficiently cautious given the absence of explicit anthropogenic variables. In the revised manuscript, we have:

- ✓ **Added a dedicated methodological discussion** on autocorrelation, persistence, and low-frequency variability, including a pre-whitening sensitivity test.
- ✓ **Completely revised the language** throughout all Sections to replace deterministic statements ("anthropogenic forcing has disrupted…") with cautious phrasing ("coincident with the industrial era", "consistent with an emerging anthropogenic influence", "cannot be uniquely attributed").
- ✓ **Added a new paragraph** in Section 6 acknowledging that our network does not include explicit anthropogenic variables and that attribution to human influence is inferential, not definitive.

Below we detail the specific changes made to the manuscript.

1. Autocorrelation, persistence, and shared low-frequency variability

We have added a pre-whitening sensitivity test (AR(1) revisualisation) and a first-difference test. The core causal pathway remained robust, confirming that our results are not artefacts of autocorrelation or shared low-frequency variability. This is now discussed *in Section 4 Methods in 4.6 and 4.7*; *Section 5* in results concretely in *5.5 and 5.6*; and *Supplementary Material S2.4 and S2.5*.

2. Revised interpretation to replace deterministic statements and structural shifts (pre-industrial vs. industrial)

We have systematically replaced over-confident language with more cautious wording.

Example, in Conclusions section *(Lines 693-697)*:

"These changes are coincident with the industrial era, during which multiple natural (reduced volcanic activity, high solar output) and anthropogenic factors (greenhouse gas emissions, land-use change, river engineering) co-vary. Because our network does not include explicit anthropogenic variables (e.g., CO_2 concentration, land cover reconstructions), the observed structural shifts should be interpreted as consistent with an emerging anthropogenic influence rather than as definitive attribution. Formal attribution would require extending the network to include those variables."

3. New cautionary paragraph added at the end of Section 6 Discussion referred to structural shifts in industrial period

We added a paragraph as introduction of section *6 Discussion (lines 499-506):*

"Before interpreting these structural shifts, two important caveats must be noted. First, our network does not include explicit anthropogenic variables (e.g., atmospheric CO_2 concentration, land-cover reconstructions, river engineering). Therefore, the observed differences between the pre-industrial and industrial DAGs reflect a composite of natural variability (changes in solar activity, volcanic frequency), non-stationarities in the flood proxy (post-1920 loss of sensitivity), and potential anthropogenic influences. Second, the industrial period also differs from the pre-industrial in terms of data quality and temporal coverage (e.g., shorter series, different proxy types). Hence, while the structural changes are consistent with an emerging anthropogenic signal, they should not be interpreted as definitive causal evidence of human-induced alteration of flood drivers. Formal attribution would require extending the network to include explicit anthropogenic forcings, which we plan for future work."

We changed the text in the original manuscript in 6.5 section:

"The structural shifts identified between pre-industrial and industrial eras (notably the weakening of solar-forced pathways and the emergence of anthropogenic drivers) mirror patterns observed in global flood attribution studies (Mallakpour & Villarini, 2015). These findings underscore the need for epoch-specific flood models, as pre-industrial analogs may poorly represent modern risk landscapes dominated by greenhouse forcing and land-use change."

By the following paragraph, *lines 535-546* in the revised manuscript:

"The comparison of pre-industrial and industrial DAGs reveals a clear reorganisation of causal pathways (e.g., weakening of solar forced precipitation and flood links, emergence of new TSI–EA connections). However, we caution against over-interpreting these changes as direct evidence of anthropogenic forcing. The industrial period differs from the pre-industrial era in multiple ways beyond anthropogenic emissions: it includes a different volcanic forcing regime (fewer large eruptions), changes in solar activity (the Modern Maximum), and, critically, alterations in the flood proxy itself due to human modifications of the floodplain after 1920 (channel corrections, embankments). Moreover, our network does not contain explicit variables representing greenhouse gas concentrations or land-use change. Therefore, the observed structural shifts should be viewed as a composite signal of all these factors. Nevertheless, the consistency of our findings with independent attribution studies that do include anthropogenic drivers (Mallakpour & Villarini, 2015; Blöschl et al., 2020) lends plausibility to the interpretation that human influence has played a role in modifying flood climate relationships. Future research should integrate explicit anthropogenic time series (e.g., CO_2 from ice cores, reconstructed land cover change) into the causal network to enable formal attribution."

We believe these revisions address the reviewer's concerns and make the interpretation appropriately cautious and transparent.

Reviewer comment:

Finally, while the Discussion section is generally informative and demonstrates appropriate hydrological and climatological background, the Conclusions section appears repetitive and occasionally somewhat

stronger in tone than warranted by the methodological limitations discussed earlier in the manuscript. I believe the Conclusions could be made more precise and better aligned in tone with the more balanced Discussion section.

Response:

We thank the reviewer for this important observation. We agree that the original Conclusions were too long, repetitive, and over-optimistic in tone, not adequately reflecting the methodological limitations discussed in the manuscript. In response, we have **completely rewritten the Conclusions section** to be more concise, precise, and aligned with the cautious tone of the Discussion.

The revised Conclusions (see below) have been shortened by approximately 40 %, repetitions have been eliminated, and the language has been moderated to better reflect the uncertainties and limitations that we now explicitly acknowledge (dating uncertainties, proxy resolution, absence of explicit anthropogenic variables, etc.). The structure has also been simplified: we removed the redundant second list ("Solar-driven atmospheric modes…") and integrated the key findings into a coherent narrative.

The revised Conclusions are now better balanced and match the more cautious tone of the Discussion. We have also ensured that the Conclusions do not introduce new claims not supported by the results.

We attach the revised Conclusions section below for the reviewer's convenience.

Revised Conclusions (as it now appears in the revised manuscript, lines 672-713)

"7 Conclusions

We have presented a causal inference framework for paleoclimate proxy time series that combines the PC-stable algorithm with bootstrap stability analysis and Granger causality validation. Applied to a multi-century dataset from the Hasli-Aare catchment (Bernese Alps), the framework identifies directed and conditionally independent relationships among solar forcing, atmospheric circulation, and summer flood variability. Our results demonstrate that constraint-based causal discovery can move beyond correlation-based attribution in paleoflood hydrology, even in the presence of dating uncertainties and proxy smoothing.

The main findings are as follows. Total solar irradiance (TSI) emerges as a robust driver of the leading mode of summer atmospheric circulation (EOF1, which corresponds to the Summer North Atlantic Oscillation). This causal link persists across the pre-industrial (1300–1849 CE) and industrial (1850–2005 CE) periods, indicating a long-term modulation of large-scale circulation by solar activity. In the pre-industrial era, TSI also exerts a detectable influence on Alpine spring precipitation (Pamj_Alps) and, through EOF1 and Pamj_Alps, on flood frequency (Flood_F1). These pathways are statistically stable under Monte Carlo perturbations of the sedimentary chronology (±20 years) and survive stricter bootstrap thresholds (0.7). Volcanic forcing (NH_Volcanic) is causally linked to Alpine summer cooling but shows only weak and indirect connections to flood variability, consistent with the transient nature of volcanic impacts.

When the same causal analysis is applied to purely instrumental data (1900–2020 CE), the core pathway TSI → EOF1 → Pamj_Alps → Flood_F1 is qualitatively reproduced, albeit with lower edge stabilities. This instrumental consistency check confirms that the inferred causal relationships are not artefacts of proxy-specific uncertainties or model assumptions.

Several structural differences emerge between the pre-industrial and industrial DAGs. The links TSI → Pamj_Alps and TSI → Flood_F1 weaken substantially in the industrial period, while new connections appear (e.g., TSI → Tjja_Alps, TSI → EA). These changes are coincident with the industrial era, during which multiple natural (reduced volcanic activity, high solar output) and anthropogenic factors (greenhouse gas emissions, land-use change, river engineering) co-vary. Because our network does ***not*** *include explicit anthropogenic variables (e.g., CO_2 concentration, land-cover reconstructions), the observed structural shifts should be interpreted as* ***consistent with*** *an emerging anthropogenic influence rather than as definitive attribution. Formal attribution would require extending the network to include those variables.*

Despite these uncertainties, the causal framework offers several advances over traditional correlation-based reconstructions. It separates direct from indirect pathways, conditions on potential confounders (e.g., volcanic forcing), and provides a quantitative measure of edge stability under dating uncertainty. The bootstrap-based thresholds (0.58 for the unconstrained DAG, 0.41 for the constrained DAG) are justified by an elbow analysis of sorted edge frequencies and are supported by sensitivity tests using stricter cut-offs.

Nevertheless, several limitations remain. The sedimentary flood proxy has an effective resolution of approximately 21 years, which smooths individual extreme events. The ±20-year dating uncertainties, while propagated through bootstrapping, still limit our ability to resolve sub-decadal causal lags. The absence of explicit anthropogenic predictors means that the industrial-era structural changes cannot be uniquely attributed to human activities. Future work should integrate higher-resolution archives (e.g., varved sediments), include reconstructions of greenhouse gases and land-use, and explore nonlinear causal discovery algorithms.

In summary, this study demonstrates that constraint-based causal inference can be fruitfully applied to multi-centennial paleoclimate and paleoflood time series. The results show that solar-driven modulation of summer atmospheric circulation has been a significant natural driver of Alpine flood variability over the pre-industrial period, but that this signal is partially masked in the industrial era. The proposed framework is reproducible and can be adapted to other mountain regions, offering a pathway towards more robust attribution of hydrological extremes in a changing climate."

We believe the revised Conclusions now match the balanced tone of the Discussion and provide a precise, non-repetitive summary of the main contributions and limitations of the study. We thank the reviewer for pushing us to improve this section.