RC1: 'Comment on egusphere-2026-303', Anonymous Referee #1, 02 Mar 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 authors investigate the potential causal linkage between climate drivers and flood frequency in the Hasli-Aare catchment from 1300 to 2020 CE. Based on different sources of climate and flood proxy records, the PC-stable algorithm combined with bootstrap stability analysis and Granger causality validation is used to detect climate-flood relationships. The results show that total solar irradiance modulates the summer atmospheric circulation, which may causally influence the flood frequency. Under modern anthropogenic forcing, these relationships are weaker than those during the preindustrial period.

In general, the structure of the manuscript is clear, and the proposed causal inference framework sounds technically promising. However, there are some concerns, as well as some more specific comments that should be addressed by the authors (please see the relevant section below), before it can be considered for publication.

Reviewer comment:

For the Introduction section, it clearly outlines the analytical framework adopted in the manuscript and provides sufficient background to understand the methodological steps that follow. However, in its current form, it reads more as a retrospective summary of the authors' previous work than as a critical positioning of the present study within the broader literature. A substantial proportion of the cited references derive from earlier publications by the same research group in the same catchment, which limits the integration of this study into the wider scientific context. Regarding the stated research gap of moving beyond correlations to identify causal mechanisms linking climate drivers and floods, this objective is relevant but not sufficiently substantiated. The Introduction does not clearly explain why correlation-based approaches are inadequate in this specific context, nor does it discuss how other studies have addressed similar questions and where key challenges persist. I encourage the authors to incorporate a broader range of external literature to demonstrate what has been achieved in this field, where traditional correlative methods face limitations, and why formal causal identification provides added value in disentangling the complexities of Alpine flood attribution.

Response:

We thank the reviewer for this constructive and insightful comment. We fully agree that the original Introduction was overly focused on our previous work and did not sufficiently position the study within the broader international literature on flood attribution and causal inference in paleoclimatology. We have therefore substantially revised the Introduction to address these concerns.

Specifically, we have:

1. **Expanded the literature review** to include key external studies on:
    - o Limitations of correlation-based approaches in climate-hydrology attribution (e.g., *Hannart et al., 2016; Ebert-Uphoff & Deng, 2012; Runge et al., 2019*).
    - o Applications of causal inference methods in Earth system science, including the PC algorithm and Granger causality (*Kretschmer et al., 2016; Nowack et al., 2020; Su et al., 2023*).
    - o Broader Alpine and European flood studies that employ or call for causal frameworks (*Blöschl et al., 2020; Wilhelm et al., 2022; Merz et al., 2021*).
2. **Explicitly justified why correlation is insufficient** in this specific Alpine context, discussing issues such as:
    - o Confounding by external forcings (e.g., solar and volcanic) that simultaneously affect multiple climate variables.
    - o Non-stationarity of flood-climate relationships across pre-industrial and industrial periods.
    - o Difficulty distinguishing direct vs. indirect pathways using correlation alone.
3. **Clarified the added value of formal causal identification** for flood attribution, highlighting how our constraint-based PC-stable algorithm with bootstrap stability analysis can:
    - o Separate spurious associations from genuine causal links.
    - o Reveal the direction of influence (e.g., TSI → SNAO → flood, not the reverse).
    - o Detect structural shifts in causal pathways under anthropogenic forcing.
4. **Reduced the emphasis on our own prior work** while retaining necessary references to the paleoflood reconstruction (Schulte et al., 2015, 2019) and SNAO definition (Peña et al., 2015) as empirical foundations. The revised Introduction now cites these as data sources rather than as the sole intellectual framework.

We believe the revised Introduction (in Main Manuscript revised) now meets the reviewer's expectations by clearly positioning our study within the wider scientific debate, justifying the need for causal inference, and demonstrating how our approach advances beyond traditional correlation-based methods

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.

Reviewer comment:

All datasets used in this study are currently described in the Supplementary Material, presumably to maintain a streamlined structure in the Methods section. While this intention is understandable, given that the analysis relies on multiple paleo-reconstructed datasets and model simulations for several key variables, I suggest moving Sections S1.1 and S1.2 into Section 2 of the main text. This would improve transparency and allow readers to more easily connect the variables introduced in the Methods with the results presented later, particularly in Section 4. As it stands, it is difficult to directly link specific outcomes to the corresponding time series without repeatedly consulting the supplementary material.

Response:

We thank the reviewer for this helpful suggestion. We fully agree that placing the detailed descriptions of flood occurrences and climate variables directly in the main text improves transparency and readability, especially given the complexity of the paleo-reconstructed datasets. Therefore, we have **moved the content of Supplementary Material Sections S1.1 and S1.2 into** ***Section 2 (Study area and previous advances: the Hasli-Aare catchment paleoflood record) and 3 (Climate Variables)*** in the Main Manuscript revised.

Specifically, we have:

- ✓ **Integrated the flood occurrence description** (originally S1.1) into a new subsection ***2 Study area and previous advances: the Hasli-Aare catchment paleoflood record***.
- ✓ **Integrated the climate variables description** (originally S1.2) into a new section ***3 Climate Variables***, including the table of variable definitions (now Table 1 in the main text) and the description of SLP grids.
- ✓ **Streamlined the Supplementary Material** by removing S1.1 and S1.2 and renumbering the remaining sections accordingly (the updated Supplementary Material will be provided as a separate file).

We believe these changes significantly improves the flow of the paper and allows readers to connect the methods directly with the results without constant reference to the supplement.

We have completely rewritten the sections 2 and 3:

2 Study area and previous advances: the Hasli-Aare catchment paleoflood record.

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

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)

Reviewer comment:

In addition, Flood_F1, which represents flood frequency and could be one of the core variables in the analysis, is only briefly described with reference to previous publications. Given its central role, more detailed information should be provided in the main text, including its temporal resolution, exact coverage period, construction method, and associated uncertainties. Similar issues apply to several other time series used in the study. This is particularly important because the title specifies 1300–2020 as the study period, whereas the end years listed in Table S2 for several series appear to fall between 1800 and 2005 and the others remain unclear. The manuscript does not clearly explain how these discrepancies in temporal coverage are handled, nor how differences in time span and resolution across datasets are reconciled within the causal analysis.

Response:

We thank the reviewer for raising this critical point. We agree that the central variables, particularly Flood_F1, require more detailed description in the main text, and that the apparent discrepancy between the title period (1300–2020) and the end years of some datasets (e.g., 2005) needs to be clearly addressed. Below we explain our revisions and clarifications.

Summary of changes made:

- ✓ **Expanded description of Flood_F1** specially in *Section 2* (Main Manuscript revised) to include:
  - o Construction method: factor analysis of geochemical elements (Zr, Ti, Sr, etc.) from delta plain sediments, calibrated against historical and instrumental flood data in *section 2.2* in Main Manuscript revised.
  - o Exact temporal coverage: 1300–2005 CE (with the post-2005 period based on instrumental flood records from the Swiss Federal Research Institute WSL) in *section 2.3* in Main Manuscript revised.
  - o Temporal resolution: annual (with ±20-year dating uncertainty for the sedimentary part, and annual precision for the instrumental period) in *section 2.4* in Main Manuscript revised.

    - Associated uncertainties: chronological (±20 years), proxy sensitivity (decreased after 1920 due to human floodplain modifications), and the Monte Carlo perturbation approach used to propagate dating errors in *section 4.2* in Main Manuscript revised.

- ✓ **Clarified the handling of the 1300–2020 period**:
    - The paleoclimate simulations (CESM-LME) and most reconstructions (TSI, volcanic, tree-ring temperature/precipitation) cover 1300–2005. The last 15 years (2005–2020) are not available for these variables. Therefore, the causal analysis focuses on the period 1300–2005, which is the common overlap of all datasets. Main Manuscript revised refers the causal modelling is performed on the 1300–2005 common interval. We have clarified this in the main text and adjusted the title. In lines 192-193: *"All series were truncated to the common period 1300–2005 CE (the overlap of all datasets) and split into a Pre-industrial (1300–1849 CE) and an Industrial (1850–2005 CE) sub-period."*, and new title: "*Causal drivers of alpine flood variability from 1300 to 2005 revealed by climate time series analysis*".
- ✓ **Added a new subsection (*3.1 in Methods*) on *"Data preprocessing and harmonisation"*** explaining:
    - How we reconciled different temporal resolutions (annual vs. decadal smoothed series).
    - How we handled different end dates: all series were truncated to the common period 1300–2005 for the causal analysis; the post-2005 flood data are shown in figures but not used in the PC-algorithm due to lack of corresponding climate driver data.
    - The ±20-year Monte Carlo perturbation to account for sedimentary dating uncertainty.
    - The use of smoothed versions (e.g., 13-year and 22-year running means) for certain variables to focus on multi-decadal variability.
- ✓ **Updated Table 3 (formerly Table S2)** to clearly indicate the exact temporal coverage of each time series and whether it was used in the causal analysis.

We believe the above revisions (expanded description of Flood_F1, clear handling of temporal discrepancies, and the new subsection on temporal harmonisation) fully address the reviewer's concerns.

Reviewer comment:

Finally, the inferred causal relationships are derived from multiple reconstructed time series based on different paleoclimate proxies and model simulations, each of which carries its own methodological and reconstruction uncertainties. These uncertainties arise from proxy calibration, dating errors, spatial representativeness, and model assumptions, and they may propagate into the causal discovery framework. At present, these datasets appear to be treated implicitly as ground truth, with limited discussion of how their uncertainties might influence the robustness of the inferred causal links. At least, the manuscript should discuss this issue more explicitly. I am also curious whether the identified causal linkages persist *when referring to the recent observational period, where instrumental data are available. Testing the proposed relationships using modern observations could provide an important consistency check and help evaluate whether the inferred mechanisms are robust.*

Response:

We thank the reviewer for this important comment. We fully acknowledge that each paleoclimate reconstruction and model simulation carries inherent uncertainties, and that treating them as ground truth is a limitation. In the revised manuscript, we have expanded the discussion of uncertainty propagation and added a new subsection in the *Discussion* (Section 6.2) to explicitly address these issues. Furthermore, we have performed an additional consistency check using purely instrumental data for the period 1900–2005, as suggested by the reviewer. Below we summarise the changes made.

Summary of changes:

- ✓ **Explicit discussion of uncertainty sources** added to the ***Methods*** **(Section 4.2)** and ***Discussion*** **(Section 6.2)** , covering:
  - o Proxy calibration errors (e.g., tree-ring temperature and precipitation reconstructions have ±0.5–1 °C and ±10–20% uncertainties, respectively).
  - o Dating uncertainties (±20 years for sedimentary flood layers, ±5–10 years for tree-ring chronologies).
  - o Spatial representativeness (e.g., the SNAO index from SLP grids may not fully capture local Alpine circulation).
  - o Model assumptions in CESM-LME (e.g., parameterisation of solar and volcanic forcing, boundary conditions).
- ✓ **Quantitative sensitivity analysis** already present in the paper (Monte Carlo perturbations, bootstrap stability, out-of-sample log-likelihood) is now explicitly linked to these uncertainty sources. We have clarified that the edge stability thresholds (58% and 41%) reflect the robustness of causal links under combined dating and proxy uncertainties.
- ✓ **New consistency check using instrumental data (1850-2005)** has been performed and is now reported in the **Discussion (Section 6.2 and 6.3)**. **Revised the Discussion** to explicitly state that while proxy uncertainties cannot be eliminated, our multi-pronged sensitivity analysis (bootstrapping, split-sample validation, instrumental consistency check) demonstrates that the main causal pathways are robust to plausible levels of uncertainty.

We believe these additions fully address the reviewer's concerns regarding uncertainty treatment and instrumental validation. The manuscript now explicitly discusses the limitations of paleo-data and quantifies the impact of uncertainties through bootstrap stability,. We are confident that this strengthens the paper's credibility and transparency.

Thank you again for this thorough and constructive review.

Reviewer comment:

Line 65, Fig. 1: It is a bit hard to find the relevant information in this figure, and I suggest separating the two panels into two figures,

Response:

We thank the reviewer and we separated in two figures:

- ✓ Figure 1. Location of study area (left) and key section AA-05 (right) in section 2.2 *The AA-05 sedimentary record: a multi proxy paleoflood archive*. Also, we added a new Table with the principal specifications of the core: *Table 1. Key characteristics of the AA 05 record (Details on lithology, geochemistry and chronology in Schulte et al. (2015))*.
- ✓ Figure 2. Multi-proxy reconstruction of summer flood variability in the Hasli Aare floodplain. Period: 1300-2005 CE in *section 2.3* . Previous reconstructions and their empirical foundation.

Reviewer comment:

Line 77, Fig. 2: It is also hard to see the context in this figure, and I suggest restructuring the layout to place the input data in the left panel and the part of the causal framework in the right panel. Furthermore, there is a lack of consistency between the figure and the Supplementary Material, as several variables in TIER 3 and TIER 5 are missing corresponding descriptions in Tables S1 and S2. Lastly, it should be "Spring Precipitation" in TIER 4.

Response:

We thank the reviewer for this careful observation. Figure 2 is corrected. We have redesigned Figure 5 (formerly Figure 2) into a two-panel layout (left: input data, right: causal framework → Tiers),

Reviewer comment:

Line 90 and the other lines: The references to tables and figures in the main text should be checked because their numbering does not match that in the supplementary materials.

Response:

We thank the reviewer for this careful observation. Checked.

Reviewer comment:

Line 168: The results provided in Table 1 are the same as in Supplementary Material 2.

Response:

We thank the reviewer for this careful observation. Checked.

Reviewer comment:

Lines 171 and 186: The choice of 0.58 as highly stable edges for the unconstrained DAG and 0.41 for the constrained DAG requires clearer statistical justification. With 100 bootstrap replicates, an edge strength of 0.58 implies absence in 42% of resamples, which may not constitute strong stability given the inherent dating uncertainties of ±20 years. Relaxation to 0.41 further raises concerns regarding the empirical support for the constrained model.

Response:

We thank the reviewer for this critical observation.

Summary of changes made:

- ✓ The thresholds (0.58 and 0.41) were not arbitrary but derived from the distribution of bootstrap edge strengths using a natural break point (elbow) analysis of the sorted edge frequencies. This is a common heuristic in causal structure learning (Scutari & Nagarajan, 2013; Friedman et al., 1999). For the unconstrained DAG, the sorted edge strengths showed a clear drop after 0.58, separating a "high stability" group from a "low stability" group. For the constrained DAG, the elbow occurred at 0.41 due to the smaller number of possible edges after blacklisting. Figures S1 and S2 in the Supplementary material.
- ✓ Therefore, we re-worded our description throughout the manuscript: instead of "highly stable" we now use "moderately stable" for edges above 0.58 (line 411 in revised main manuscript), and *"The constrained DAG resulted in a lower significance threshold for edge retention*" (line 434 in the revised main manuscript) for edges between 0.41 and 0.58.
- ✓ A sensitivity analysis is added (Supplementary Material S2.3. Sensitivity to bootstrap threshold) showing that the main causal pathway (TSI → EOF1 → Pamj_Alps → Flood_F1) persists even at stricter thresholds (0.7 or 0.8, , in Supplementary Material S2.3. Sensitivity to bootstrap threshold, lines 173-180 and new *Table S4: Edge stabilities (bootstrap frequencies) for key directed edges at different thresholds*).
- ✓ The lower threshold for the constrained DAG (0.41) is justified because the blacklisting of non-physical edges greatly reduces the candidate edge space, so even moderately low bootstrap support still indicates the most plausible link given physical knowledge. We have also added references to similar practices in the Earth system science literature (Kretschmer et al., 2016; Runge et al., 2019). We believe these clarifications and additional analyses address the reviewer's concern.
- ✓ Emphasize that bootstrap stability is only one measure of robustness. Our confidence in the causal links also comes from: (i) Granger causality tests (temporal precedence), (ii) split-sample validation across periods, (iii) physical plausibility. Even edges with moderate bootstrap stability (e.g., 0.58) are supported by these independent lines of evidence.
- ✓ Reported results of sensitivity analysis of backlist constraints in new section in *Supplementary Material S2.6* (Lines 203-213 and *Table S7. Comparison of network properties under different blacklist constraints*).

Reviewer comment:

Line 205: For section 4.5, what is the bootstrap stability threshold for the DAG? While differences in DAG topology between pre-industrial and industrial periods demonstrate structural evolution in the inferred network, further caution is needed before interpreting these changes to anthropogenic forcing since no explicit anthropogenic variables were included in the network.

Response:

We thank the reviewer for this careful observation. We will address two points: (1) clarifying the bootstrap stability threshold used in Section 4.5, and (2) adding a cautionary note about attributing structural changes specifically to anthropogenic forcing given the absence of explicit anthropogenic variables (e.g., greenhouse gas concentrations, land-use change) in the network.

Summary of changes:

- ✓ **Clarify the bootstrap stability threshold for the industrial-period DAG**. We now explicitly state in Section 4.5 that the same threshold (0.41 for the constrained DAG) was used for both pre-industrial and industrial networks. This ensures comparability. We also note that a sensitivity analysis using a higher threshold (0.7 or 0.8) does not alter the main conclusions (in *Supplementary Material S2.3. Sensitivity to bootstrap threshold*).

- ✓ **Cautious language on anthropogenic attribution:** We fully agree that, without explicit anthropogenic variables (e.g., CO_2, land-use change), interpreting the observed structural shifts as definitive evidence of human influence is premature. Therefore, we have revised the language throughout *Section 5 Results* and *6 Discussion* to replace deterministic statements ("anthropogenic forcing disrupted…") with more cautious wording, such as "consistent with the hypothesis of anthropogenic influence", "coincident with the industrial era", and "could reflect a combination of natural and human factors". We have also added a new paragraph (at the end of Section 6.1, lines 535-546) explicitly acknowledging that our network lacks direct anthropogenic predictors and that the industrial period differs from the pre-industrial in multiple aspects (volcanic forcing, solar activity, proxy sensitivity). This paragraph notes that formal attribution would require the inclusion of explicit anthropogenic time series, which we plan for future research: *"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 (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."*

- ✓ **Update the relevant text** in **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 believe these revisions fully satisfy the reviewer's request.

Reviewer comment:

Lines 277-282 and Lines 290-294: These parts appear to contain repetitive content.

Response:

We thank the reviewer for pointing out this redundancy. Upon re-examining the manuscript, we confirm that the two passages indeed contain overlapping statements regarding the interpretation of causal pathways and the role of solar forcing. We have now consolidated these sentences into a single, clearer statement and removed the repetition.

Specifically:

- **Lines 277–282** (originally in the Discussion) discussed the cascade of influences from external forcings to atmospheric circulation to flood risk.
- **Lines 290–294** (also in the Discussion) repeated a similar idea with slightly different wording.

We have now **merged** these two passages into a single concise paragraph (new lines 653-659 in the revised manuscript) and deleted the duplicate. The revised text is provided below. We have also checked the rest of the manuscript for other repetitions and found none.

Thank you again for helping us improve the clarity and conciseness of the paper.

Before (repetitive):

[Lines 277–282, original]
"Taken together, these pathways suggest a cascade of influences: external forcings modulate atmospheric circulation (tier 1 → tier 2), which in turn governs both moisture delivery and thermal conditions (tier 2→ tier 3), ultimately shaping flood risk (tier 3 → tier 4). By enforcing physical constraints via blacklists and tiering, the causal graph avoids spurious feedback and yields an interpretable model that aligns with mechanistic understanding."

[Lines 290–294, original]
"This comprehensive methodology enhances our understanding of the interplay between external forcings, climate components and their collective impact on flood risk. The discussion highlights key findings and explores their broader implications for flood management and climate adaptation strategies, emphasizing the importance of integrating global climate drivers and regional responses into predictive models (Wilby & Harris, 2006; Kundzewicz et al., 2013)."

After (consolidated, no repetition):

[New lines 653-659]
"Taken together, these pathways suggest a cascade of influences: external forcings modulate atmospheric circulation (tier 1→tier 2), which in turn governs both moisture delivery and thermal conditions (tier 2→ tier 3), ultimately shaping flood risk (tier 3 → tier 4). By enforcing physical constraints via blacklists and tiering, the causal graph avoids spurious feedback and yields an interpretable model that aligns with mechanistic understanding. This comprehensive approach enhances our ability to integrate global climate drivers and regional responses into predictive flood risk models, as also advocated by Wilby & Harris (2006) and Kundzewicz et al. (2013). The following sections explore the broader implications for flood management and climate adaptation."