Reply to Reviewer 1

We would like to thank both referees for their constructive feedback. We will gladly work out a revision of the manuscript along the received comments.

Accordingly, we would like to make the following major changes to the manuscript:

- The discussion in this article should focus on the spatio-temporal flood patterns in the Alps using Generalized Additive Models (GAMs).
- Detailed information on the statistical analysis (Sections 2, 4.1, and 4.2) should be moved to the supplementary material.
- We will provide more detailed information on the sediment data and its specific characteristics (catchment area, seasonality, etc.).
- We will revise the structure of the article to better guide interested readers from various disciplines, while highlighting the article's contribution to the still-open hydroclimatic questions: (a) whether flood events will increase or decrease in a warmer climate, (b) what specific regional flood patterns occur over time.

Below are more detailed responses to the general comments of the report. All the comments can and will be addressed in a revision.

General comments

Comment. The analysis extend the work of Wilhelm et al. (2022) by using GAMM models to investigate the space-time dynamic of flood frequency in the Alps in the last 1450 years. I liked reading the paper although I have found the theoretical treatment at a level higher than my knowledge, so consider my comments on that part as the view of a typical reader of NHESS (i.e., not a mathematician).

Reply.

> *Thank you very much for the careful review of our paper and for expressing your interest in reading it. We would like to emphasize again the large potential of long flood records from Alpine lake sediments. In our work, we aim to present a robust statistical approach that takes into account the unique characteristics of such sediment data. We understand the concern that the focus on statistical details could be reduced in the paper. Therefore, we would like to move details to the supplementary material so that the results do not dilute the main hydrological questions.*

Comment. The manuscript in its current form gives the impression that the available dataset is primarily being used as a vehicle to showcase a statistical method, rather than that the methodological choices are guided by well-defined hydrological questions.

Reply.

Thank you for this valuable comment. We will revise the motivation and framing accordingly. In particular, we will describe the dataset in more detail and move the statistical details to the supplementary material. Based on our analysis, we will focus more explicitly on the spatio-temporal variability of floods in the Alpine region and on the open hydro-climatological question of whether floods increase or decrease under further climate warming.

Comment. The data description is insufficiently detailed to support several of the interpretations, particularly those related to the apparent irrelevance of catchment characteristics (see the detailed comments below).

Reply.

We agree that the data description should be expanded in the manuscript. Although the sediment data have been described in more detail in Wilhelm et al. (2022), we will add a dedicated description of the lake and catchment characteristics used in our analysis. This will make the manuscript more self-contained and will better support the interpretation of the catchment-related results.

Comment. Several methodological passages are not sufficiently clear, at least to me, particularly with respect to notation, the meaning of individual model components, and the role of the penalty terms (see the detailed comments below).

Reply.

In the revised manuscript, we will place greater emphasis on interpretation and readability, while preserving the necessary statistical rigor. We will also move technical details that are not essential for understanding the hydrological results to the supplementary material, so that the main text can better guide readers through the modelling choices and their implications.

Comment. Finally, the discussion of the results also needs to be strengthened. Some claims are insufficiently substantiated, potential biases and artefacts in Figure 7 are not fully addressed, and the relative importance of temporal, spatial, and space-time effects is not discussed clearly enough.

Reply.

We will strengthen the discussion further by better connecting the statistical treatment of the data, the spatio-temporal model, and the interpretation of flood activity in the Alps over the last 1450 years. We will also place the results more clearly in the context of available evidence, including historical flood records where appropriate.

Detailed comments

Comment. Title: I wonder whether the wording “Flood Risk” is misleading. Indeed the analysis is on the flood frequency, not touching on the consequences. Probably “Flood Hazard” is a better wording.

Reply.

Thank you for this important remark. This is indeed misleading, and we will revise the title accordingly. We will further use flood activity instead of flood risk.

Comment. Line 8: what does "susceptibility" mean?

Reply.

> *This should mark potential clusters of floods. We will change the term "susceptibility" to "occurrence" to avoid any confusion.*

Comment. Line 9: since climate variables are not used as covariates, how can the model be used for hazard assessment under changing climate conditions?

Reply.

> *We have not yet used any climate variables because the available datasets have insufficient temporal and spatial resolution. This can be addressed in the future as soon as climate data are available. Then, the model could be used for risk assessment under changing climate conditions. Our explanation may not be clear enough, so we will rephrase the sentence as: "This provides a robust tool for future analyses ..."*

Comment. Lines 21–27: relevant literature is cited here. While Blöschl et al. (2017, 2019) are certainly relevant, I was surprised that Blöschl et al. (2020, `https://doi.org/10.1038/s41586-020-2478-3`) is not considered, since it could be partially used as a benchmark study (the analysis spans "only" 500 years and is at a larger spatial scale).

Reply.

> *We will add this paper to our discussion as well.*

Comment. Line 83: maybe I would already state here what typical link functions are (e.g., the logarithm). I was puzzled for a while on what could g be while reading the method section.

Reply.

> *In Section 2, where we introduce the GAMMs, we will include examples of link functions, such as the identity link function for Gaussian responses, the logit link function for binary responses, and the logarithmic link function for count data.*

Comment. Line 113: I see the importance of penalizing too flexible models. However I am a bit lost in where equation (6) (or 7) relates to the model (equation 2). The same issue (maybe just mine) applies to the penalty term at line 133.

Reply.

> *We agree that the connection between the model for the conditional expectation in (2) and the penalty terms should be made clearer. We will revise the text to make this connection explicit. The penalties enter at the estimation stage, and are applied to unknown quantities in the model. For the smooth functions, they act on the basis coefficients introduced in (6), and for the random effects, they act on the corresponding random-effect coefficients introduced around line 133. These penalties are added to the model deviance when the model is fitted. The smooth penalties prevent fitted curves or surfaces from becoming too irregular, while the random-effect penalty prevents lake-specific or group-specific deviations from becoming unnecessarily large. We will also clarify that (7) is a shrinkage version of (6), meaning that, besides controlling the irregularity of a smooth term, it allows the estimation procedure to shrink the whole term towards a negligible effect when it is not supported by the data.*

Comment. Equation (9): please define all terms, including sigma and gamma.

Reply.

> *We thank you for pointing this out. We will make sure to clarify that γ is the semivariogram and σ^2 is the marginal variance at a spatial location. In particular, we will connect equation (9) to the discussion in Section 4.2. Moreover, we consider the covariance between responses at two spatial locations and two time epochs.*

Comment. Line 189: I would state already what other approach is used instead.

Reply.

> *We will include it in the revised manuscript.*

Comment. Lines 190–200: the database is described here. I would suggest to give more quantitative details. What ranges do catchment area and elevation span? In the results you will show that these characteristics do not matter, is it because they are all very small catchments at high elevation? I think having an idea on the catchments involved could be useful in the interpretation of the results by the readers.

Reply.

> *The sediment data and the characteristics of the catchment were described in Wilhelm et al. (2022). We will also expand on the description of the data in our paper to support our interpretation of the lake catchment characteristics.*

Comment. Equation (14): where does this variance come from? From the theory in Selch et al. (2018)?

Reply.

> *The variance in (14) is the standard variance of the Negative Binomial distribution under the parametrisation used in the manuscript, while the work of Selch et al. (2018) motivates modeling the increments of the counts, rather than the occurrence of a flood for each year. We will clarify this at the beginning of Section 4.1.*

Comment. Line 235: what does "originate from" mean?

Reply.

> *In equation (17), we specify the additive predictor of the model, that is, the component that combines the offset, intercept, random effects, and covariate effects. Through the logarithmic link function, this additive predictor determines the conditional expectation of the count increments. We agree that the current wording "originates from" is unclear, and we will replace it with a more precise formulation.*

Comment. Equation (17): please define all terms. I had an hard time to retrieve their meaning.

Reply.

> *We thank the reviewer for this helpful comment. Indeed, equation (17) contains several components whose meaning should be made explicit where the model is introduced. We will add a short explanation defining each term. This will include the response being modelled, the role of the time-window offset, the baseline intercept, the lake-specific and season-specific random effects, the temporal, spatial, and space-time smooth terms, and the smooth terms for altitude, catchment area, and maximum catchment altitude.*

Comment. Lines 237–238: I do not understand the sentence, sorry. Maybe it could be expanded.

Reply.

> *The role of the offset is to account for the length of the interval over which floods are counted. Here, an offset means a fixed term included in the predictor, rather than an effect estimated from the data. Since the response is the number of floods in a time interval, longer intervals are expected to contain larger counts. Then, including the time interval as an offset makes the expected count proportional to the length of that interval. After this adjustment, the remaining exponential term can be interpreted as an annual flood occurrence rate, or annual expected number of floods. This allows the spatial, temporal, spatio-temporal, and catchment-related components of the model to be interpreted as effects on the count increment after accounting for the size of the time window. We will point this out clearly in the revised manuscript.*

Comment. Line 262: here the penalty matrix is finally revealed. Maybe a hint on what P could be could have been given before, when P is first introduced.

Reply.

> *We agree that the connection should be made earlier. We will clarify that each penalised component of the model has its own penalty matrix. The matrix introduced around line 262 is associated specifically with the spatial random effects. It corresponds to the block of the overall penalty structure that penalises the lake-specific random intercepts with spatial dependence. We will add this explanation when penalty matrices are first introduced, so that Section 4.2 is easier to follow.*

Comment. Line 275: this is true, but what is the size of the catchments in here? And the elevations? More intense convective precipitation is expected to occur at low elevation, for instance.

Reply.

> *The lake sediment records derive from various sites that differ in catchment scale (1–4600 km^2) and elevation (231–2620 m a.s.l.). We will add this information in the paper.*

Comment. Line 284: what is the default number of basis functions in `mgcv`.

Reply.

> *Your comment gives us the chance to improve the clarity of the text. We believe that the statement "mgcv default number of basis functions" is too vague, because the default basis dimension depends on the type of smoother and on the dimension of the covariate space. We will explicitly report the basis dimensions used for each smooth term in equation (17).*

Comment. Lines 293–294: is penalization to 0 the result of the analysis? Can we see it in Table 1? I see edf=0 for ca, caalt and alt, but not for season. And since Table 1 seems to indicate that the 4 terms are not significant (p-value), should not they be removed from the model, like I would do in a linear regression?

Reply.

> *Yes, the shrinkage of some terms to a negligible effect is a result of the analysis, but we agree that our wording should distinguish this result from the non-significance of*

the season effect. In Table 1, the smooth terms for catchment area, maximum catchment altitude, and altitude have effective degrees of freedom close to zero. This indicates that the penalisation has shrunk these effects to essentially negligible contributions, which is consistent with their large p-values. The season effect is different. Its effective degrees of freedom are not close to zero, but its p-value indicates that it is not statistically significant. We will revise the manuscript to make this distinction clear.

Regarding the removal of terms, the penalised GAMM framework differs from a standard linear-regression setting. The shrinkage penalty acts as a safeguard against overfitting by reducing unsupported smooth terms to negligible influence during estimation. Therefore, smooth terms that are shrunk to negligible effects do not need to be manually removed for the purposes of model interpretation. We will clarify this point in the revised text.

Comment. Table 1: the terms in the table should be defined. E.g., what is edf? And Ref.df? What do they mean?

Reply.

We agree that Table 1 should be self-contained. Moving forward, we will add definitions of the quantities reported in the table caption. In particular, we will define edf as the effective degrees of freedom of a fitted smooth or random-effect term, which indicates how much flexibility the term uses after penalisation. Values close to zero indicate that the term has been shrunk to a negligible effect, while larger values indicate a more flexible fitted effect. We will define Ref.df as the reference degrees of freedom used for the approximate significance test reported in the table. We will also define the remaining entries in the table.

Comment. Lines 301–307: are these lake-specific deviations on top of the smooth spatial deviation discussed later? If so, should not they be discussed later?

Reply.

The lake-specific random effects are best interpreted as deviations from the overall constant intercept. They capture persistent lake-level differences in flood occurrence among the observed sediment records. By contrast, the smooth spatial term describes a continuous regional pattern across the Alpine domain. We will revise the text around lines 301–307 to clarify that Figure 3 shows lake-specific random-intercept effects, which are shifts from the overall estimated intercept of the model.

Comment. Line 310–325: this is an interesting result. It seems coherent with Blöschl et al. (2020, `https://doi.org/10.1038/s41586-020-2478-3`) for the last part of the series.

Reply.

We will include this in the discussion.

Comment. Lines 326–329: Figure 5 puzzles me. It seems not to be consistent with Figure 4. Of course Figure 4 contains only the temporal effect. Is Figure 5 also affected by the fact that not all lakes have series going to the early period? In other words, if we count few event in the past because we have data from few lakes, is not the figure misleading because of a systematic bias?

Reply.

We thank the reviewer for this important observation. Figure 5 is intended as a qualitative descriptive check, not as a direct quantitative counterpart of Figure 4, and we will clarify this in the revised manuscript. In particular, Figure 4 shows the estimated temporal effect from the full model, after accounting for the other components, while Figure 5 is based on the aggregated observed counts. We agree that, if raw aggregated counts are shown, the figure can be affected by the number of lake records contributing to each time window. We will address this by normalising the descriptive counts by the number of available lake records in each 20-year window.

Comment. Lines 330–346: Figure 4 ranges from 0.7 to 1.4, while Figures 6 and, more so, 7 present even larger ranges. Does it mean that spatial variability is dominant with respect to time? And that space-time clusters are even more important?

Reply.

The ranges shown in Figures 4, 6, and 7 should not be interpreted as directly comparable measures of relative importance. These figures display different partial effects of the fitted model, shown separately after exponentiation. Therefore, the larger visual range of the spatial or space-time effects does not by itself imply that these components are dominant over the temporal effect. We will revise the text to clarify that the figures illustrate different model components and should not be read as a formal comparison of effect magnitude.

Comment. Besides, the locations where in Figure 7 the darker colours occur are far from the data (the little black points). Does not this indicate that there may be artifacts? Would these artifacts have an effect on the results in Figure 4 and 6 too?

Reply.

We agree that regions far from the observed lake locations in Figure 7 should be interpreted with caution. At each selected time point, Figure 7 shows a smooth spatial surface fitted from irregularly distributed lake records, so values far from the black points may reflect extrapolation and may be more prone to artifacts. We will clarify this limitation in the revised manuscript. This mainly affects the interpretation of the spatial maps, rather than invalidating the temporal effect in Figure 4 or the broad spatial effect in Figure 6.

Comment. Also, for the discussion, would not it be useful to have a figure like Figure 7 where all smooth functions are added (and maybe also the global average term)? This would clarify whether spatial or temporal variability dominates.

Reply.

We agree that a figure showing the combined fitted surface could help interpret the model, and we will consider adding such a figure.

Comment. Figures 9 and 10: the axes labels are not nicely written.

Reply.

Thank you for pointing this out. This will be fixed in the revised manuscript.

Comment. Line 377: I would use "long-term flood frequency dynamics" instead of "flood dynamics", which are usually referring to the dynamic of individual events.

Reply.

We will adopt this wording in the revised manuscript.

Comment. Line 379: the irrelevance of catchment area may be due to the fact that the catchments analyzed here are not so different in size. Please provide the numbers. Same comment regarding the remarks at lines 421–422.

Reply.

The catchment scale differs from 1–4600 km^2*. We will add this information in the paper.*

Comment. Line 406: would not other climate indices than temperature be useful covariates here?

Reply.

We have not yet found any suitable proxy data with sufficient temporal resolution. That would be a goal for the future.