Reply to Reviewer 2

We would like to thank both reviewers 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 article is clearly too statistically oriented for the NHESS journal. For example, the GAMM framework is presented before the data, as if the data were merely an illustration of the statistical method. I do not think this is appropriate for NHESS: the authors should start with the data and the scientific problem, and then present the model they implemented to address the questions they aim to tackle. Furthermore, many terms in the methods section are highly mathematical, such as "compact domain", "covariate vector X" (without specifying what X is, since the data are presented in the following section), "row vector of covariates W" (same remark), "design vector for random effects Z", etc.

Reply.

> *We thank the reviewer for his comments. We believe that this topic is of interest for NHESS readers and that interdisciplinary approaches are becoming more important for such complex hydrological questions. But we understand that the statistical background is highly mathematical and the details might be too specific for NHESS community. Therefore, we will revise and reframe the paper. Specifically, we will focus on spatio-temporal flood activity. We will move the detailed statistical background to the supplementary material.*

Comment. I got lost in the description of the model and its interpretation. It seems to me that the quantity of interest is $N_s(t)$, the number of floods at time t and location s, but what is actually modelled is not $N_s(t)$ directly but the increments in the number of floods between t and $t + \tau$ ($Y_s(t)$). If I understood correctly, the conditional expectation of $Y_s(t)$ is modelled by a

GAMM as $\exp(\eta_s(t))$, given by equation (17). Then, all the results are presented through the different components of equation (17). However, it is not clear to me how equation (17) is linked to $N_s(t)$, which is the quantity of interest. Consequently, it is unclear how the results should be interpreted in terms of flood probability.

Reply.

> *Our intention is to model the expected number of new floods over a time interval. Since the cumulative flood history is obtained by adding these increments over time, modelling the expected increments provides a way to describe the evolution of the cumulative flood count. Moreover, the model output is an expected flood count, or an annual flood occurrence rate once the interval length has been accounted for. Therefore, when we refer to higher or lower flood likelihood, we mean that the model expects more or fewer floods over a given period. We will revise the text to make these points clearer.*

Comment. Model validation in Section 5.2 is difficult to understand for non-statisticians. It would be preferable to validate the different results (Figs. 3-7) against empirical estimates.

Reply.

> *We agree that Section 5.2 should be made more accessible to non-statistical readers, and we will revise the explanation of the diagnostic plots accordingly. However, we would keep the simulation-based residual diagnostics as the main validation tool, because the purpose of this section is to assess whether the model assumptions are reasonable. Descriptive empirical summaries are useful for visual interpretation, but they do not provide a check for assumption violations. We will clarify this distinction in the revised manuscript and, where useful, refer to descriptive comparisons as interpretive aids.*

Comment. The article is a step forward compared to Wilhelm et al. (2022) in terms of statistical modelling, but it is nowhere clearly stated what new scientific insights are gained relative to Wilhelm et al. (2022). Since NHESS is not an applied statistics journal, I do not think that focusing mainly on the statistical model and its technical aspects is appropriate for this journal.

Reply.

> *We will emphasize the hydrological insights gained from jointly modelling the lake records in space and time, including the identification of regional flood patterns, temporal changes in flood activity, and the role of lake-specific and catchment-related characteristics. In particular, we will position our scientific contribution relative to Wilhelm et al. (2022) clearly. We will move technical material to the supplementary material so that the main text focuses more clearly on the new hydroclimatic interpretation.*

Comment. In summary, the current format of the article is not suitable for NHESS. I encourage the authors to substantially revise the manuscript in order to: (i) make it understandable to non-statisticians and better connect the model to the scientific problem and flood risk itself - the reader needs much more guidance throughout the paper; (ii) better validate the results against the data and empirical estimates, which would also help readers understand the model; and (iii) more clearly discuss what new knowledge is gained compared to Wilhelm et al. (2022).

Reply.

We thank the reviewer for this clear summary. We will substantially revise the structure and presentation of the manuscript along these lines. In particular, we will improve the guidance for non-statistical readers, better connect the model to the hydrological questions, strengthen the interpretation of the results, and clarify the new scientific insights obtained from the spatio-temporal analysis of the Alpine lake records.

Additional comments

Comment. L192-199: I think a table summarizing the characteristics of the catchments used as covariates in the article is needed, even if this information is already available in Wilhelm et al. (2022), because: (i) I believe not all columns from Extended Data Table 1 of Wilhelm et al. are used here; and (ii) readers should not need to consult another article to understand your model (i.e., the covariates) and therefore your results.

Reply.

The revised manuscript should be self-contained with respect to the covariates used in the model. Thus, we will add a table summarising the lake and catchment characteristics used in our analysis.

Comment. equation (17): Please make the link with equation (2) explicit by clearly defining g, W_{st}, Z_s, f_i, and X_{st}. Many variables appearing in this equation are undefined; please explain each term. The notations (alt_s), (ca_s), etc. have never been defined previously. Why do you have both ($b(s)$) and ($f(s)$)? What is the difference between them?

Reply.

We agree that the link between the general GAMM formulation in equation (2) and the applied model in equation (17) should be made explicit. We will add a comprehensive paragraph defining all elements of equation (17) and mapping them to the notation of equation (2), including the link function, the covariates, the random-effect terms, and the smooth functions. We will also define the variables used in the applied model, such as altitude, catchment area, and maximum catchment altitude.

Comment. All maps except Fig. 1: it is difficult to identify where the main risk areas are located within the Alps. Please add country borders.

Reply.

We will add a more detailed map that includes borders.

Comment. Fig. 3: please explain more specifically what it means to have $\exp(u_s) = 3$ compared with $\exp(u_s) = 1$, for example in terms of flood risk.

Reply.

The lake-specific random effects represent persistent deviations of individual lakes from the global intercept, after accounting for the other model components. Since Figure 3 shows these effects on the multiplicative scale, a value of 3 means that, for the same time period and covariates, the lake-specific component triples the expected number of floods in the interval compared with a lake whose value is 1. We will add a concrete example to clarify Figure 3.

Comment. Figs. 4-7: similarly, please explain the values in terms of flood probability (and not only in terms of "more" or "less" risky).

Reply.

> *To make the interpretation of Figures 4–7 clearer, in the revised manuscript, we will specify that the plotted values represent multiplicative effects on the expected flood count, or annual flood occurrence rate after accounting for the interval length. Thus, they should not be read directly as flood probabilities. When we refer to flood probability or likelihood, we will clarify that this is expressed through changes in the expected number of floods.*

Comment. Validation: Section 5.2 is again too strongly oriented towards mathematicians. Figs. 8-10 are difficult for non-statisticians to interpret. In contrast, Figs. 3-7 seem to be interpretable in terms of flood probability and are therefore easier to understand. Would it be possible to provide empirical estimates (at station locations) of the quantities shown in these figures in order to assess the goodness-of-fit of the model?

Reply.

> *As mentioned previously, we will rewrite the diagnostic section for a broader readership. We will consider adding empirical summaries at the lake locations as visual aids for interpreting the fitted effects. However, we will keep the simulation-based residual diagnostics as the main goodness-of-fit assessment, because they are designed to detect assumption violations that descriptive empirical summaries alone cannot rule out.*