Reviewer 1

This paper uses generalized additive models (GAMs) to compare pre- and post-dam construction flows as a function of rainfall, temperature and glacier mass changes in 74 high mountain catchments. The approach enables to reconstruct reservoir operations, as well as its hydro-climatic drivers, in situation where inflows and outflows are not readily available (which is often the case). The paper is really interesting, the method is a smart way to tackle a challenging puzzle, and there is a feeling that it is likely novel… but this is incumbent on authors to demonstrate, both by explaining the novelty in more detail, and by validating the method! My main remarks are as follows:
Reply: *Thank you very much for this valuable and constructive feedback, which we highly appreciate. Please find our responses to the detailed comments below.*

1) the novelty needs to be explicitly and precisely stated on the introduction.
Reply: *Thank you very much for highlighting the need to more specifically describe the novelty of our study. The novelty is twofold, as specified in the introduction. First, we 'propose a statistical three-step approach for reservoir signal reconstruction in catchments where reservoir outflow but no inflow time series are available'. Second, we 'shed light on spatial variations in reservoir regulation signals and their relationship to catchment characteristics'. 'The combination of the proposed reservoir signal reconstruction approach with functional clustering allows us to provide insights into how reservoir regulation varies spatially in the Alps and to which degree these variations are related to catchment characteristics.'*

2) the absence of a formal validation is a real issue that should (and can!) be addressed.
Reply: *Thank you for pointing out the need of a formal validation. We expanded both the validation of the GAM model and compared the extracted reservoir regulation signals to an alternative signal derived from regional reservoir storage curves provided by the Swiss Federal Office of Energy (for more detail see our response below).*

3) the methodology is creative, but the explanations are sometimes vague, and needs to be better justified and documented in several places.
Reply: *Thank you very much for indicating the need for clarification. We carefully revised and expanded the methods section, by (1) adding an illustration of the workflow (Figure 3); (2) moving the technical information from the introduction to the methods section; and (3) providing additional details on the different working steps, in particular model evaluation.*

4) results need to be presented step by step. They are really interesting, but authors need to be more rigorous in presenting how they got to them.
Reply: *We present the results related to model evaluation in the Methods section together with detailed figure descriptions, while we present the results related to the reservoir signal variation analysis in the results section. We consolidated the results section by redesigning Figures 7 and 8 following recommendations by Reviewer 2.*

It might take a substantial revision for authors to address the comments above. Revisions are important because this work has the potential to be a nice paper. I now develop the four points above.

1) Several points need to be clarified in the introduction:

=> What is the exact scope of the paper? Is the type of data authors base themselves on widely available in a range of cases (not just the Alps)?
Reply: *Thank you for highlighting the need to better describe the scope of the paper. We specified in the introduction that: 'This approach can be used to reconstruct reservoir operation signals in catchments where streamflow and climate data are available for a period before and after a known date of reservoir construction. Such information is more widely available than reservoir in- and outflow measurements, which means that the approach is applicable in different regions around the globe where streamflow observations and information on reservoir construction dates are available. Here, we apply the approach to extract reservoir signals from observed time series of 74 catchments in the Central Alps.'*

=> Why are GAM the appropriate methodology here? Why not use machine learning methods for instance?
Reply: *We agree that other methods could potentially be used for predicting natural streamflow. We aimed to identify a relatively simple and parsimonious model using a small number of explanatory variables. Using machine learning approaches would only make sense if many different explanatory variables were used for model fitting, which is not the case here. We added a short discussion of the advantages of GAMs compared to simpler and more complex model alternatives to the introduction: 'The main advantage of GAMs is that cubic spline modeling offers flexibility for each covariate and goes beyond a restrictive linear regression framework, while the additive structure among covariates remains simple. This balance between non-parametric modeling and a simple additive link facilitates the interpretation of the contribution of each explanatory variable. Still, other regressions techniques (neural networks, random forest and other ML algorithms) could replace our GAM approach in the scheme displayed in Figure 3. Keeping in mind that our training period can be short (a few decades) at some locations, this lack of a large training dataset may also limit the application of fully data-driven machine learning techniques.'*

=> A simple search of "generalized additive model reservoir" on Scopus returns 120 results. Can authors ascertain no similar attempt has been tried in the past? What have been GAM uses in the literature around hydrological and reservoir modelling?
Reply: *GAMs are versatile and flexible models for a range of prediction applications and have been previously used in the context of reservoirs, e.g. to predict reservoir sedimentation loads or reservoir water quality. We acknowledged these alternative uses of GAMs in the context of reservoirs in the introduction: 'In this study, the response variable yt corresponds to streamflow time series in mm/d (units). Alternatively, GAMs have in the context of reservoirs also been used to predict other variables than streamflow such eutrophication levels (Catherine et al. 2010) or downstream water temperatures (Coleman et al. 2021). ' However, we are unaware of any study that has used GAMs to infer the reservoir operation signal from observed streamflow time series covering a period before and after reservoir construction.*

2) Because there is no inflow and outflow data for the catchment authors study, they do not validate the approach using basins for which the post-dam construction reservoir inflows and outflows are known. But there are many mountain reservoirs in other regions (e.g., the USA) for which inflows and outflows, as well as the other data authors use in the Alps. This would provide

formal validation. Note they can then choose to split this paper into two papers (i) one that presents and validates the method, (ii) one that applies it to Alpine catchments to derive new knowledge (which is what they have here). I note here that the confidential dataset shared by authors does not address this concern about validation.

Reply: *We agree that some sort of direct validation on observed reservoir storage changes is desirable. Observed inflow and outflow data are hardly available and naturalized inflow is often simulated. Therefore, we looked for an alternative validation strategy. The Swiss Federal Office of Energy provides weekly reservoir storage estimates aggregated over a larger region (i.e. canton). We used these regional storage estimates to compute seasonal changes in regional storage. We then used the regional storage change curve derived for the region Valais to evaluate the reservoir signal extracted using the GAM for a catchment located in that region (see Figure 1). The regulation signal estimated for the Rhône catchment using the GAM approach compares really well to the signal derived from observed regional reservoir storage data. We added this evaluation based on observed data to the manuscript.*

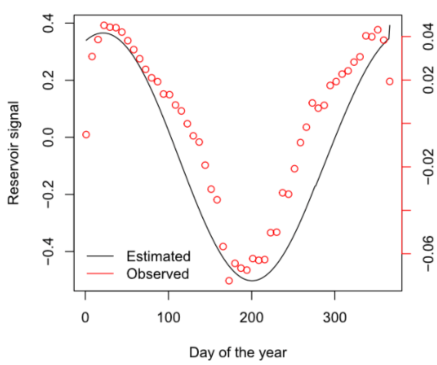

Figure 1: Observed reservoir storage change curve for the region Valais derived from regional reservoir storage data provided by the Swiss Federal Office of Energy compared to the estimated reservoir regulation signal for the catchment Rhône (Porte-du-Scex), which is located in the Valais region.

3) In order of appearance, my remarks are as follows:

Performance of the pre-dam model: with 74 catchments, it would only be normal that the GAMs model would work more or less well across the sample. The methodology section (lines 126-130 in particular) does not explain what performance measures are used. Related question: do authors investigate the influence of the length of the pre-development record on model quality? Please include a full list (maybe in supplementary material) of the catchments, including relevant characteristics (e.g., flow record length, year reservoir went online, catchment surface, ratio of reservoir storage over natural annual flow).

Reply: *Thank you for highlighting that a broader goodness-of-fit assessment was required. We specified in the methods section that: 'We assess the model's performance by comparing observed with predicted streamflow values and by computing a range of different performance metrics including the Kling-Gupta (KGE) and Nash-Sutcliffe efficiencies (NSE) (Gupta et al. 2009, Nash 1970), volumetric efficiency (Criss et al. 2008), mean absolute error (MAE), root mean squared error (RMSE), and percent bias (PB).' The results of this analysis are presented in the new Table 1. In addition, we added a full list of all catchments including their catchment characteristics to the* *appendix (Table 1A): Country, river, measurement station, record length, catchment area, mean elevation, and start year of reservoir operation.*

Authors give no indication whether the possibility of an extended filling period for the reservoirs was considered. Or did they assume this would be negligible? If so, please justify explicitly.
Reply: *We specified that with 'reservoir construction date', we mean the 'date when the reservoir went into operation.'*

Section 2.2: the method enables one to derive time series of inflows / storage / release. Authors choose a final signal in terms of difference outflows-inflows that is creative and innovative, and could be a key contribution of the paper. Yet it is not justified, especially in relation to the literature on the role of reservoir in hydrological models that authors claim to contribute to. This leads to several important questions:

i) How do the chosen signals relate literature trying to specify reservoir outflows as a function of inflows, storage, and reservoir purpose?
Reply: *Thank you for highlighting the need to discuss similarities and differences to other types of signals described in the literature. We specified that: 'This resulting signal indicates how much water is stored in and released from reservoirs in which season (i.e. day of the year). These reservoir-storage-seasonality signals take a reservoir perspective and provide information on storage in addition to releases, but not on inflow. Therefore, they are distinct from the signals extracted through other approaches, e.g. simulated water releases (Coerver et al. 2018); spectral differences between in- and outflows highlighting the time scales most affected by reservoir regulation (White et al. 2005, Shiau et al. 2014); or water storage and release policies, which define release decisions as a percent deviation from long term mean inflow (Turner et al. 2021).'*

ii) Why not derive relationships between inflows and outflows once inflows are derived? That is more standard.
Reply: *Thank you for this suggestion. Our approach does not provide information on reservoir inflows. Instead, it provides information on when and how much water is released from and stored in reservoirs. We clarified this in the introduction by writing: 'This resulting signal indicates how much water is stored in and released from reservoirs in which season (i.e. day of the year). These reservoir-storage-seasonality signals take a reservoir perspective and provide information on storage in addition to releases, but not on inflow.'*

iii) How do we know smoothing actually is needed? And how do we know how to parameterise the smoothing technique?
Reply: *We compared the performance of the GAM driven with smoothed covariates with the performance of a GAM driven with raw (unsmoothed) covariates. This comparison showed that model performance can be improved by using smoothed instead of raw data, which is why we decided to use the smoothed covariates for our model. A comparison of observed natural flow with simulated flow derived using a GAM with smoothed covariates and a GAM with raw covariates is shown in Figure 2, which indicates that simulated values are closer to the observed values when using the GAM with the smoothed covariates. This visual impression is confirmed when computing and comparing the NSE and KGE performance metrics for the two model types across all 74*

catchments (Figure 3 in this response to the reviewers). In almost all catchments, the GAM with smoothed covariates performs better both in terms of KGE and NSE.

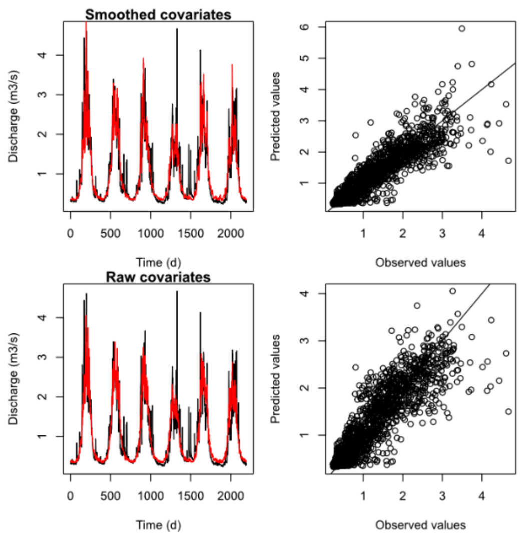

Figure 2: Comparison of simulated and observed natural flow for an example catchment where the simulated flow is derived with a GAM fitted on the smoothed covariates (model used in the study, upper panel) and a GAM fitted on the raw covariates (lower panel).

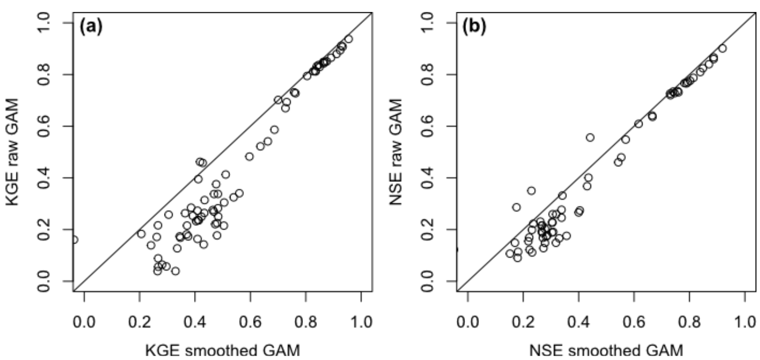

Figure 3: Goodness-of-fit ((a) KGE and (b) NSE) of observed and simulated flow derived using two types of models: (1) GAM with smoothed covariates and (2) GAM with raw covariates.

4) The results section is very interesting in highlighting a category of high-elevation hydropower reservoirs (but authors should comment in the discussion how new this is). But it only reports on what happens once the flow signal is derived, whereas it would be great to see results for the fitting of the pre-dam GAM model, as this is key to getting quality results. Authors should report in detail on goodness of fit, relationship between number of years of record and quality of the results, etc.
Reply: *Thank you for highlighting the need for a more in-depth performance assessment. We provide results of the fitting of the pre-dam ('natural') GAM model in the methods section. In Figure 4, we compare observed natural flows to simulated natural flows of an example catchment. We extended the performance analysis to all catchments by computing a range of different goodness-of-fit statistics (Nash-Sutcliffe efficiency, Kling-Gupta efficiency, percent bias, volumetric efficiency, mean absolute error, and the root mean squared error) for all basins and display them in the new Table 1. We also added a new figure to the appendix showing the relationship of model*

performance (in terms of KGE) to the length of the record available for model fitting, catchment area, and elevation. We described the results of this additional goodness-of-fit analysis as follows: 'Model performance is independent of the length of the record available to fit the GAM but depends on catchment area and elevation (Figure B1). The best performance is achieved in large and high-elevation catchments, while performance is worst in small and low-elevation catchments.'

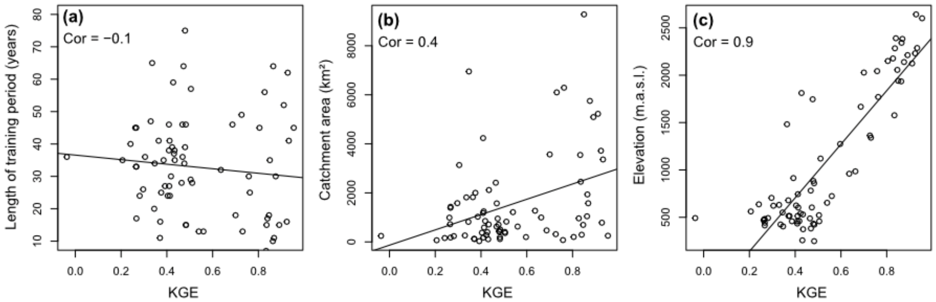

Figure 4: Kling-Gupta efficiency vs. length of training period, catchment area, and elevation.

Another remark on results at lines 161-163: this is a key point that seems difficult to make without taking a long look at reservoirs' storage capacities. Reservoir with small regulation capacity (ratio of live storage to average annual flow) will naturally have weaker signals.
Reply: *We expanded the description of the results and added a subpanel to the new Figure 8 (d) showing the distribution of reservoir storage capacities across catchments belonging to clusters 1 and 2. The catchments with the larger storage capacities are those with the stronger seasonal redistribution.*

One last remark concerning results is that it would be great to highlight one or two known water supply reservoirs (for irrigation or any other uses), and show how their method enables quantifying downstream needs. That's not strictly necessary but it would show that the proposed method is versatile (which it is), instead of being a one-trick pony that only identified high-elevation single use hydropower reservoir.
Reply: *Thank you very much for this suggestion of using the extracted regulation signals to estimate downstream water demand. We think that the reservoir signals cannot be used as direct estimates of water demand because part of the water demand might be covered by other sources of water such as groundwater storage or precipitation.*

Finally, below are also some minor remarks:

Lines 47-51: You talk about a two step setup and then lose the reader a bit by not making both steps explicit. I'd advise doing that.
Reply: *Thank you for highlighting the need for clarification. We now talk about a 'three-step' approach and specified that: 'In a first step, the approach fits a GAM to streamflow observations representing natural pre-reservoir conditions using precipitation, temperature, day of the year, and glacier mass balance changes as covariates. In a second step, this GAM is applied to covariates derived for the regulated post-dam period to predict natural streamflow for this regulated period. In a last step, the reservoir regulation signal is reconstructed by subtracting the predicted 'natural signal' from the observed regulated signal.'*

Line 56: by convention, please introduce an equation with ":" Several times in the paper please replace "see, e.g." with "see e.g.,".
Reply: *We added ':' before equations and replaced 'see, e.g.' by 'see e.g.'.*

Line 113: please replace "(GAM) (Hastie" by "(GAM; Hastie" Figure 3.c: the x-axis says this is the early 1950s whereas the legend says early 1910s. Which is it? Also, please fit the limits on the x-axis to the beginning and ending of the record you are plotting (also valid in Figure 4).
Reply: *We corrected the text formatting and corrected the figure caption.*

Figure 4: please enlarge the figure.
Reply: *We enlarged the figure.*

"Glaciermelt": consider writing "glacier melt" instead.
Reply: *We replaced 'glaciermelt' with 'glacier melt'.*

Line 203-207: yes, but it would be great to try to nail the scope in the introduction instead of just speculating in the discussion. How applicable is this data to other places? That can be in principle deduced from the methodological setup alone.
Reply: *We specified the data-requirements for the application of the approach in the discussion: 'The GAM-approach proposed here can be used to reconstruct reservoir operation signals in other parts of the world, given that streamflow and climate data are available for a period before and after a known date of reservoir construction.' We also stated in the introduction that: 'This approach can be used to reconstruct reservoir operation signals in catchments where streamflow and climate data are available for a period before and after a known date of reservoir construction. Here, we apply the approach to extract reservoir signals from observed time series of 74 catchments in the Central Alps.'*