

Reviewer 3

I liked reading this paper which is nicely and clearly written. The results are not surprising but anyway useful in providing some quantitative statements about the effects of the influence of reservoirs/lakes (and area) on the ratio between hourly and daily peaks. I have just some comments that I hope will be useful to improve the paper through a minor revision. More generally I think that:

Reply: *Thank you very much for acknowledging the value of our work and your suggestions on how to further improve the manuscript.*

- Since the ratio between daily and sub-daily peaks does not refer to individual events, but to estimated quantiles, this should be stated clearly in the title/abstract of the paper. Why hasn't the analysis focused on individual events? With the data available this would have been possible and, in principle, would have allowed an even clearer attribution of the dampening to the reservoir/lakes, I think. Some clarification on this choice is needed.

Reply: *Thank you for highlighting the need to justify this choice. We indeed compared estimated flood quantiles rather than individual peaks. We added the following justification for this choice to the methods section:*

'To compare daily and sub-daily flood peaks, we work with flood estimates, that is flood quantiles estimated for specific return periods. We used such estimates instead of focusing on individual events as flood events detected in one time series are not necessarily detected in the other series.'

Furthermore, we adjusted both the title and abstract to better reflect this choice by highlighting that we are talking about 'estimated flood peaks'.

- The methodology section should provide some more details on the methods and the reasons for choosing them. For example I have not understood why the analysis of sensitivity of peak flows to the catchment area is performed through a log-linear model "without an intercept" (see detailed comments below). Also, it is unclear to me how Figure 5b and Figure 6 are obtained. Some math (not too much) would maybe help.

Reply: *Thank you for highlighting the need to clarify some methodological choices like the choice of a linear model without an intercept, the quantification of increase in node purity (shown in Figure 5b), and the partial prediction of the peak ratio (shown in Figure 6). We specified these choices in the methods section as follows:*

- *intercept: 'We chose a linear model with no intercept to force the regression line through the origin (0,0) because a catchment with a catchment area of 0 should also have a 10-yearly flow of 0.'*

- *increase in node impurity: 'To identify the most important characteristics, we used two metrics derived from the random forest model to quantify variable importance: The increase in mean squared error (% INC MSE) and the increase in node purity. [...]. The increase in node purity reflects a characteristic's importance in improving the decision tree's structure by quantifying the total decrease in impurity as quantified by the Gini*

index from splits using a characteristic (Liaw and Wiener 2002).'

- partial prediction of the peak ratio: 'Additionally, we use partial dependence plots (Friedman 2001 and Greenwell 2016) to estimate the marginal effect of the most important variables for predicting the D/H ratio. These plots help visualize the relationship between a subset of the predictors and the response while accounting for the average effect of the other predictors in the model. In doing so, they reveal the marginal relationship between the predicted variable and each predictor and can be used to assess the type of relation across the value range of each variable. We use these partial dependence plots to identify when the contributing area percentage has a strong influence on the D/H ratio.'

- As also stressed by another Reviewer, while the analyses done to compare the “area” and the “contributing area percentage” effects are certainly valuable and most important, it is strange that the other attributes, which from Figure 5 appear not that less-important, are then neglected in the analyses and discussions.

Reply: *We added the following justification for our focus on the two most important variables to the results section: 'Our further analyses focus on the influence of the two most important characteristics, namely, contributing area percentage and catchment area, because they seem to be the most important predictors of the D/H ratio related to the 10-yearly flood peak.' Furthermore, we expanded the discussion on the other two predictors considered: 'While catchment area and contributing area percentage are the most important predictors of the D/H ratio, other predictors such as the biogeographical region, a metric for catchment similarity in terms of climate, geology, and vegetation, and other catchment characteristics such as permeability, an indicator for flashiness, only have limited predictive power (Fig. 5). This highlights that existing estimation approaches deriving instantaneous peak flows from daily data using catchment area and other catchment characteristics can profit from additionally including information on water body influence.'*

Detailed comments:

Lines 85-90: The Walensee does not have a small difference in contributing areas between the upstream and downstream gauges. Why then not including other Swiss cases in the Case-Study analysis?

Reply: *While we would have loved to include additional or alternative Swiss case studies, we did not find any other cases fulfilling the criterion of having up- and downstream gauges with sub-daily streamflow records.*

Line 119: I don't understand the sentence “we excluded lakes which are part of other lake catchments”. Even the following example does not clarify the point.

Reply: *Thank you for highlighting the need for clarification. We rewrote the sentence to 'Since we are interested in the total lake-influenced catchment area, we only included the most downstream lake within a catchment.'*

Line 144: The ratio between daily and sub-daily peaks does not refer to individual events, but to estimated quantiles. This was not so clear in the previous part of the paper. Maybe the title or, at least, the abstract should include this information. Besides, it would be interesting to know in how many years the daily maxima and the subdaily maxima do not correspond to the same event.

Thank you for highlighting the need to highlight the focus on estimated quantiles. We indeed compared estimated flood quantiles rather than individual peaks. We added the following justification for this choice to the methods section:

‘To compare daily and sub-daily flood peaks, we work with flood estimates, that is flood quantiles estimated for specific return periods. We used such estimates instead of focusing on individual events as flood events detected in one time series are not necessarily detected in the other series.’

Furthermore, we adjusted both the title and abstract to better reflect this choice.

Line 160: How is “impurity” defined?

Reply: *We specified that ‘impurity’ was defined using the Gini index: ‘The increase in node purity reflects a characteristic’s importance in improving the decision tree’s structure by quantifying the total decrease in impurity as quantified by the Gini index from splits using a characteristic (Liaw and Wiener 2002).’*

Line 162: From the description here, I cannot grasp how the “partial dependence plots” work.

Reply: *Thank you for highlighting the need for clarification. PDPs help visualize the relationship between a subset of the predictors and the response while accounting for the average effect of the other predictors in the model. We specified this in the manuscript.*

Line 174: Why is it necessary to fit log-linear models without an intercept? Is the equation then $Q = A^{\beta}$? Why should Q be equal to 1 when the area is 1? The property declared in the following section, “The slope shows by how many % points extreme flows increase when the catchment size increases by 1 %”, also holds for the model $Q = \alpha \cdot A^{\beta}$, i.e., a log-linear model with the intercept. Maybe adding the equation would be beneficial. I am not a fan of a too mathematical treatment in hydrology papers but sometimes math helps the reader.

Reply: *By setting the intercept to 0, we force the model through the origin (0,0), thereby assuming that a catchment with 0 km² area should also have 0 flow. We justified this choice in the manuscript as follows: ‘We chose a linear model with no intercept to force the regression line through the origin (0,0) because a catchment with a catchment area of 0 should also have a 10-yearly flow of 0. The slope of this regression model shows by how many % points extreme flows increase when the catchment size increases by 1 %.’*

Figure 2: Is there a reason for scaling logarithmically the y-axis (flow estimates)?

Besides, data could be added here with a plotting position formula.

Reply: *We logarithmically scaled the y-axis to limit its vertical extent. We used a theoretical instead of an empirical distribution to derive flow estimates for specific return periods.*

Figure 4: Maybe 4 points corresponding to the downstream gauges in the Case-Study analysis could be added here as a reference. This would demonstrate that the two analyses are consistent (even though not geographically).

Reply: *Thank you for this suggestion. While we think that this is a nice idea, we refrained from doing this because only one of the case study catchments is located in Switzerland (see Figure 1 in the manuscript).*

Lines 222-225: This result is interesting. Even though not dominant, also the biogeo_region and permeability attributes seem important. Why are they neglected in the following analyses?

Reply: *We focused the subsequent analysis on the two most important predictors (contributing area percentage and catchment area) in order to disentangle the influence of reservoirs from the one of catchment area.*

Figure 6: As said above, I do not understand how these diagrams are obtained.

Reply: *Thank you for highlighting the need for clarification. PDPs help visualize the relationship between a subset of the predictors and the response while accounting for the average effect of the other predictors in the model. We specified this in the manuscript. We used the R-package pdp by Greenwell to construct the plots.*

Lines 239-245: I wonder if your experimental setup is “fair”. In Figure 7 the brown line is calibrated and the light-blue/green one is in extrapolation (if I understand it well). Wouldn't be better to calibrate the model on some no-water-body influenced catchments and then apply it to both non-influenced and influenced catchments not used in calibration? I am confident that the result would be similar but not affected by the calibration issue. Or, if the Authors want to demonstrate that the “contributing area percentage” is necessary, wouldn't it be better to use all catchments for calibrating the model without that attribute? Probably, in that case, also the brown distribution would be biased but in the overestimation range.

Reply: *It is correct that the random forest model used to predict the results shown in Figure 7 was trained on weakly water-body influenced catchments only. What we are trying to show is that doing so will allow us to predict the D/H ratio in these catchments, while it will fail to do so in regulated catchments. This highlights that information on water body influence is crucial to predict the D/H ratio in strongly regulated catchments. We clarified the purpose of this analysis in the methods section of the manuscript: ‘This analysis aims to highlight the importance of water-body influence information for explaining the D/H ratio in regulated catchments.’ Calibrating the model on a subsample of weakly water-body influenced catchments only would have substantially reduced the amount of data available for model fitting, which is why we would like to avoid this.*

Random forest models avoid overfitting by measuring prediction error on an out-of-bag sample that is sampled randomly in each iteration of the model fitting process.

Figure 8: Indeed from the figure I assume that the model used is $Q = A^\beta$ (the lines meet at the point (1, 1)). It is not clear to me why this should be preferred to $Q = \alpha A^\beta$. Some clarification is needed in the method part.

Reply: *As explained above, we force the model through the origin (0,0) by setting the intercept to 0, thereby assuming that a catchment with 0 km² area should also have 0 flow. We justified this choice in the manuscript as follows: ‘We chose a linear model with no intercept to force the regression line through the origin (0,0) because a catchment with a catchment area of 0 should also have a 10-yearly flow of 0. The slope of this regression model shows by how many % points extreme flows increase when the catchment size increases by 1 %.’*