

Reply to review by Lieke Melsen

We thank Dr. Lieke Melsen her careful read to the manuscript and insightful suggestions.

After reading carefully the comments of both the reviewers, in order to address some of their most substantial comments, there will be some major changes in the revised version of the manuscript. These changes will affect most significantly the data analysis part of the paper (section 3), and they are summarized hereafter:

- We will extend the selection of hydrological signatures in order to have a more comprehensive picture of the hydrological behavior of the catchments.
- We will select only signatures that are not correlated (reducing substantially their number) for the subsequent analyses.
- We will select a reduced number of catchment and meteorological indices in order to reduce the problem of correlated features.
- All the correlation analysis will be based on the Spearman's rank score since, as pointed out by Dr. Melsen, the usage of Pearson may not be adequate.
- Following this decision, we will remove the regression analysis (table 3) since (1), as stated by the Anonymous Referee #1, "it does not add any new knowledge to the literature occupies a lot of space" and because (2) the usage of linear regression (that looks for linear dependencies in the data) is not coherent with the choice of basing the analysis on the Spearman's rank score (which accounts also for nonlinear correlations).

Below, we answer in detail the various comments, and illustrate our intended approach to address them in the revised version. The original comments of the reviewer are reported in *black and italics*, our replies in blue.

Dal Molin et al. investigated, through regression analysis, which indices have explanatory power for streamflow response. Based on the insights gained from the regression analysis, different spatial configurations were implemented in a hydrological model. Although I find the work flow elegant, starting from process-understanding and translating that to the spatial configuration of the model, I have some problems / concerns with the regression set-up.

Major

My main concerns are all related to the regression-part of the study.

1. *It is unclear how the indices, on which regression was applied, were selected. There is plenty of literature around on indices and signatures, which could guide indices selection, but I don't see any justification in the text for the choices made. Check for instance: Addor et al., A Ranking of Hydrological Signatures Based on Their Predictability in Space, WRR, 2018*

Knoben et al., A Quantitative Hydrological Climate Classification Evaluated With Independent Streamflow Data, WRR, 2018

We thank the reviewer for the references she provided. We will use them to better contextualize, and potentially expand, our selection.

The signatures used in this work were chosen in order to represent a wide variety of hydrograph characteristics. There are, for instance, two signatures designed to represent the long-term water balance (average streamflow and runoff coefficient), two signatures to capture the “responsiveness” of the hydrograph (baseflow index and flashiness index), and one (the half streamflow period) that is designed to understand seasonality effects, like the ones related to the snow dynamics.

The indices used (here by indices we mean catchment and meteorological indices) were selected in order to have a large group of possible influence factors to start the analysis with. Catchment characteristics capture topography, soil properties, land use, and geology while meteorological characteristics take into account precipitation and potential evapotranspiration.

We will improve the description of the indices, including the motivation behind their selection and integrate or modify the list of indices also according to what has been proposed by the other reviewer.

2. *Since the choice of the indices is not well justified, I am worried about their mutual correlation. Many indices can describe the same signal. Therefore, please provide the correlation among the indices themselves. This might lead to the insight that you need fewer or different ones.*

Mutual correlation, as pointed out by the reviewer, is a potential criticality of this study. Indeed there are multiple indices that describe the same signal (or very similar signals) and this possibility was considered in the design of this study. We decided to keep them all because we didn’t want to restrict a priori the space of possible influence factors. However, we will check, in the revised version, for mutual correlation between indices before carrying out the correlation with hydrological signatures.

We will add tables with the correlations between indices themselves in the revised version to make clearer the choices made in section 3.2.2.

3. *It was rightly mentioned that correlation does not mean causality. It was claimed (not only in the methods, but also in the conclusions) that this study accounts for that by only selecting the indices that have a causal relation, based on expert judgement. I do, however, not recognize the expert judgement in the selection of significant indices, and this actually directly relates to my point 2. Right now, the selection seems to be made based on the mutual correlation of the indices – so the mutual correlation was investigated! – but I don’t see any process-reasoning (the expert-judgement) that can justify the selected indices, and that justifies the claim that there is really causality.*

The analysis of the correlations was only the starting point of the process of selection of the indices. Starting from that, we then discarded the indices that are either redundant (e.g. average slope vs. fraction of steep areas) or accidental correlations (e.g. using elevation instead of precipitation, line 18 page 7). This step was essential to move from mere correlation to causality and it involved “expert judgment” in order to prune reasonably the list of indices. An example of

the “expert judgment” process is illustrated in paragraph 3.2.2, page 10, lines 1 to 10, where we showed how the indices were selected for the mean streamflow signature.

We will improve our text (especially paragraph 3.2.2) in order to highlight the role of “expert judgment” in the process of indices selection. We will also try to reduce the dimensionality of the problem selecting a sub-set of signatures, indices, and characteristics before doing the correlation analysis.

4. *I disagree with the conclusion of the authors that there is no need to look for nonlinearity in the correlation, based on the results in the table. The authors rightly state that only few correlations that are statistically significant based on Spearman are not significant with Pearson, but how do the authors explain the opposite effect? Quite some correlations are significant with Pearson but not with Spearman, is this a Type 2 error in the Pearson test? That could have consequences, for example, aridity was significantly correlated with BFI for Pearson, but not Spearman, and based on ‘expert judgement’ included in the regression.*

The reviewer is right pointing out the possibility of type 2 error for some correlations, motivated by discording conclusions between Spearman and Pearson correlation. This may be due to the fact that the Pearson correlation was calculated neglecting the assumptions behind it (e.g. normality of the data) that may not be respected in this case.

We will remove from the revised version the usage of Pearson correlation and use only Spearman’s rank since it relaxes some of the assumptions made by Pearson and it also detects nonlinear correlations.

5. *1 of the 3 points of the guidelines for modelling based on the regression was not based on the regression at all, namely the conclusion that the presence of snow is relevant. Please include a snow-related indicator in the regression to support this conclusion (based on expert judgement we can expect this, of course).*

The reviewer is right saying that it is not clear that “the conclusion that the presence of snow is relevant” was motivated by the regression analysis. This is due to the fact that we didn’t explain earlier in the text that the “half streamflow period” signature and the plots in figure 5 were used to show the presence of seasonality in the streamflow dynamics. In particular, there are some subcatchments that reach their peaks of streamflow in different periods of the year.

This seasonality is not due to different patterns in precipitation or PET (as shown in figure 5) and correlates well with the elevation (higher subcatchments reach half of their streamflow later in the year). These two points made us think that this seasonality is due to snow dynamics and that, therefore, the model should take them into account; higher catchments are subjected to snow that is then released in the streamflow (as snowmelt) later in the year if compared with rain-dominated subcatchments.

We should mention earlier in the paper the possibility of snow playing a role in this catchment and that therefore the signature and figure 5 are used to capture this effect.

6. *It depends a bit on the definition of model building, but the title and the text might give the impression that the model structure itself was adapted with the insights in the regression, while it was basically the model implementation (accounting for HRU’s or not) that was adapted.*

It is true: it depends on the definition of model building. For us it incorporates all the decisions taken in order to have an hydrological model for the Thur catchment. In particular:

- How to spatially divide the inputs
- How to divide the catchments in HRUs
- Structure of the single HRUs

All these points were considered in the construction of the hydrological model and were informed by the regression analysis. The last point (structure of the bucket model) was also considered in the procedure of building the hydrological model but was not discussed in this paper because it was already done in previous studies and for sake of brevity.

We will make more clear what do we mean for model building in the introduction of the paper.

Minor

Section 3.2.1, the catchments are sort of grouped based on their stream flow response, but this is not used any further in the analysis. Consider to just briefly describe their response, or to use the grouping later to explain results (in that case, also display the groups in the figure).

Section 3.2.1 describes the signatures in the catchments without relating them with the indices. We agree with the reviewer that the subdivision done here is no more used in the paper and it was done only for convenience when describing the signatures. We will remove this division since it is not relevant for the rest of the paper.

In the same section, it seems unnecessarily complicated to use combinations of signatures to determine how flashy catchments are; a flow duration curve can generally provide quite some insight on this already (slope of flow duration curve also frequently used signature)

We acknowledge the possibility of using other signatures to describe the behavior of a catchment but we used, among the others, baseflow index and the flashiness index because believe that they are more interpretable and they can be related to dynamics represented by the model; the BFI, for example, can be linked to the separation between quick and slow flow that is a process that is present also in the hydrological model. The selection of the signatures will be revised in the new version of the paper also according to the previous points and to the other reviewer.

Provide an overview of the indices and their abbreviations, or include their full name more often in text / tables / figures, because now it requires quite some work from the reader to fully understand all sentences and figures (and a lot of going back to the methods).

We understand that the usage of abbreviations may complicate the reading of the paper but, on the other side, their usage helps avoiding misunderstandings that may happen when calling the same index with different names. The full name of the indices is provided in table 2 and in section 3.1.1 and it will be reported in the caption of the figures when not reported in the figure itself.

A large number of figures is dedicated to showing the signature-values, which is not of direct relevance. I would be interested to see a figure that displays the HRU's.

We acknowledge that a figure representing the HRUs used in M2 and M3 is missing but it can be deducted from figure 1 (plot "d" for M2 and "c" for M3) since the HRUs were constructed aggregating

some classes (for example, for M2, one HRU is composed by the orange part and the other by the rest of the catchment). We think that the figures dedicated to showing the signature-values are important for this study to show the high variability present in the catchment response.

For the landscape characteristics, it is not mentioned in the methods-section (3.1.1) that you consider fractional area. Please clarify there, as I was wondering how you would apply regression on nominal values, until I found out in the results that you considered frac. area.

We will clarify it in the text.

The sentence ‘optimizing the parameter of the posterior distribution’ (l.11, p13) can give the impression that you minimized e.g. variance (describing distribution), please consider reformulation.

We need to clarify it in the text. The actual meaning is “optimizing the parameters of the hydrological model and of the error model (refer to section 4.1.1 and 4.1.2) in order to find the ones that maximize the posterior distribution”

Although overall written well and clear, some language editing seems required, for example “The average value oscillates of about ..”, (but I’m not a native either).

We will do our best to improve it.

Overall, I appreciate the intent of the study and the modelling-part seems well designed (except for my question at point 6 which remained unclear), but I do believe the regression-part requires substantial revision, related to the selection of indices (more embedded in literature and account for snow) and to justify the use of the word ‘causality’. Given that the work-flow is largely set-up, I think the authors should be able to incorporate this.