Author's response

Changes in the paper

The paper has been subject to major revision in order to address the comments of the reviewers. The most significant changes concern the signatures analysis, where we introduced an extended selection of signatures, followed by an approach for selecting significant signatures, and a new model experiment (M0), aimed at demonstrating causes for differences in streamflow seasonality. A summary of the changes is presented below, followed by the individual responses to the reviewers.

Changes in text

SECTION	DESCRIPTION
1	Minor changes to keep consistency with the rest of the paper.
3.1 and 3.2	Completely restructured to address the comments of the reviewers: • The list of signatures and indices considered was extended. • Correlation analysis based on Spearman instead of Pearson correlation to account for nonlinearities. • A pre-selection of signatures and indices based on correlation analysis and expert judgment was done to avoid spurious correlations in the subsequent analyses. • Climatic and catchment indices that drive streamflow variability were identified using correlation analysis and expert judgment.
3.3	Adapted to the new findings.
4.1.1	Completely restructured to address the concerns of the reviewers about clarity.
4.1.4	Minor changes to keep consistency with the rest of the paper.
4.1.5	Completely restructured to address the concerns of the reviewers about clarity. New model M0 included to test the importance of snow-related processes.
4.2.1	Minor changes to include M0.
4.2.2	Major changes due to the new group of signatures considered.
4.2.3	Minor changes to keep consistency with the new findings of section 4.
5	Minor changes to keep consistency with the rest of the paper.
6	Major changes to address the concerns of the reviewers.

Changes in Figures

The following figures were eliminated or modified (note that the numbers refer to the old version of the paper)

FIGURE	DESCRIPTION
1	Panels (a), (b), and (c) modified to address the concerns of the reviewer.
2	Eliminated.
3	Eliminated.
4	Eliminated.
5	Eliminated.
7	Added M0.
8	Eliminated.
9	Eliminated.
10	Eliminated.

The following figures were created (note that the numbers refer to the new version of the paper)

FIGURE	DESCRIPTION
2	Internal correlation between the streamflow signatures.
3	Internal correlation between the climatic indices.
4	Internal correlation between the catchment characteristics.
5	Correlation between the selected streamflow signatures and the selected climatic and catchment indices.
8	Influence of the model structure on the representation of the mean half streamflow day.
9	Ability of the models of representing the signatures.
10	Ability of the hydrological models of representing the signature duration of low-flow events.

Changes in Tables

The following tables were eliminated or modified (note that the numbers refer to the old version of the paper)

Table	DESCRIPTION
1	Some columns were eliminated because they were included in table 4 of the new version of the paper.
2	Eliminated.
3	Eliminated.
A1	Minor changes to address the comments of the reviewers.
A2	Minor changes to address the comments of the reviewers.
A3	Minor changes to address the comments of the reviewers.

The following tables were created (note that the numbers refer to the new version of the paper)

FIGURE	DESCRIPTION
2	Values of the streamflow signatures.
3	Values of the climatic indices.
4	Values of the subcatchment characteristics.

Reply to review by Anonymous Referee #1

We thank the reviewer for his/her careful read to the manuscript and insightful suggestions. Below, we answer in detail the various comments, and illustrate how we have addressed them in the revised version. The original comments of the reviewer are reported in *black and italics*, our replies in blue and the changes to the paper in **green and bold.**

Note that, were not indicated, the numbers of figure/tables/lines/pages refer to the old version of the paper.

The authors propose to infer the structure of a hydrological model based on landscape and process characteristics (signatures) of the catchment. In the first of a two-stage process different landscape and catchment characteristics are compared to different streamflow signatures to identify the most important controls on runoff formation. In the second step this information is used "as an inspiration for model structure design" (p17, l. 32) as the authors put it.

Inferring structure from function (or vice versa) is at the core of hydrological model building and subject to numerous studies. The topic is hence highly relevant for the hydrological community. The manuscript is well structured and well-written. Accordingly the manuscript is suitable for a publication on HESS. However, I cannot recommend publishing to current version of the manuscript due to several major points:

1. *The purpose of the modelling exercise is not clear. Model requirements for flood forecasting are e.g. totally different from model requirements to simulate climate change. The relevant signatures, temporal and spatial model discretisation, model evaluation metrics and also the degree of model conceptualisation differ accordingly. Please specify more clearly the purpose of you modelling study. Otherwise it is not possible to evaluate the study meaningfully.*
   In the introduction, we already specified (lines 6-10, page 3) the main objectives of the study, which in summary consist of proposing a model building strategy which starts from an analysis of the data, which provide a basis for motivating the various model decisions.
   As the reviewer noted, the purpose of the model itself has remained unclear. Although one of the main objectives of models is making predictions to address some practical issues, here we do not have such an immediate applied objective. The model exercise is mainly an instrument to help understand and interpret catchment scale processes. In particular, we are interested in identifying the landscape properties and associated processes that dominate catchment response, and that mostly influence the observed spatial variability in streamflow behavior, as characterized by the set of suggested signatures.
   The reviewer is right in the fact that aspects such as "signatures, temporal and spatial model discretization, model evaluation metrics and also the degree of model conceptualization differ accordingly".
   We have addressed this comment specifying (line 5, page 3 in the new version of the paper) that the objective is explaining the observed spatial diversity of streamflow characteristics, with the minimum possible complexity, while trying to maintain a process based interpretation.

2. *I consider the selection, evaluation and identification of landscape characteristics as fairly weak due to a number of different reasons:*
    a. *The authors provide no information about why certain characteristics were selected (and why others were not). Catchment characteristics (or signatures) can only provide information on the underlying processes if they have some kind of diagnostic potential or causal relationship. It is clear that these relationships are often unknown and difficult to obtain; nevertheless the selection of appropriate characteristics is vital for the identification of underlying processes and mechanisms. I miss a clear and elaborate description on the selection of catchment descriptors and on their expected diagnostic potential (both in space and time): E.g. why or how can the different land cover ratios or aspects help to derive information on hydrological processes? Are the same characteristics suitable for all catchments (independent of size, altitude, geology)? Please also comment on the importance of the time step e.g. you calculated the flashiness index based on daily streamflow data, although you state that streamflow can change two orders of magnitude in a few hours (p. 3 l. 18). If this is true please explain why you consider a daily-data based flashiness index as a meaningful variable? Please do also explain why you think that "half streamflow period" is a suitable parameter to discriminate to importance of snow. I expect that there are much simpler and more meaningful variables such as temperature and rain, temperature sums or snow data itself to describe the importance of snow. The results in Fig 7-9 also show that streamflow, runoff coefficient and half streamflow period are pretty identically in all cases. Do you consider them being suitable signatures? Please also provide more signature papers in the introduction as the number of up to date references is small.*

       The reviewer correctly points out that we "miss a clear and elaborate description on the selection of catchment descriptors and on their expected diagnostic potential (both in space and time)". Catchment characteristics can affect the hydrological cycle: vegetation characteristics, for example, are typically assumed to affect evaporation, soil characteristics are typically assumed to influence the partitioning of water between retention and runoff. In general, we tried to select a broad class of characteristics, to be as inclusive as possible. However, it is also true that these characteristics can be represented through a large class of indices, and in order to reduce the size of the problem, some choices had to be made.

       We have complemented in section 3.1.1 of the new paper the selection of catchment characteristics with their expected diagnostic potential. We have also motivated some of our decisions based on how other models have dealt with similar issues.

       We used a daily data resolution, and this choice clearly affects some of the signatures. As the reviewer points out, the flashiness index is one of such signatures. The values of the flashiness index reduces with increasing time step due to a smoothing effect. In this paper, we did not experiment with varying data resolution, as it was outside our scope.

       We have commented (line 25-30, page 4 of the new version of the paper) about this choice relating it to findings of previous modeling studies.

We experimented with several signatures to account for the effect of snow on streamflow seasonality, and we ended up selecting the "half streamflow period". The reason is that 1) this signature was used in previous publications to quantify streamflow seasonality, and we did not want to invent our own signature if something was already existing, and 2) this signature captured well the difference between streamflow regimes, because we have seen (figure 5) that all the catchments receive similar precipitation input (in terms of monthly variability) but the snow-affected ones show the peaks during late spring/beginning of summer while the rain dominated ones show their peaks during the winter and the spring.

Figures 7-9 show that all the model configurations represent well the yearly streamflow, the runoff coefficient, and the half streamflow period and this is a result of our study that is also coherent with our assumption that only distributing the inputs (precipitation, PET, and temperature) is sufficient in order to have a model that captures the water balance and the snow dynamics.

We have extended the list of signatures considered in this study and used correlation analysis between the signatures to select only the not redundant ones (major changes in section 3 of the new version of the paper).

b. *In your study you included several (fairly easy to derive) landscape characteristics that are obviously highly correlated and describe in great detail how you identify and select appropriate ones based on regression and correlation. In my opinion a rather trivial part which does not add any new knowledge to the literature occupies a lot of space. I hence suggest shortening and streamlining the entire section. If you want to derive structure from function than the first goal must be to derive a (comprehensive) matrix of uncorrelated catchment characteristics that have some kind of diagnostic potential. In my opinion this should be the source of the story and not a result.*

Section 3 of the new version of the paper has been completely restructured to address (also) this comment. The lists of meteorological and climatic indices have been reduced before evaluating the correlation with the streamflow signatures; the regression analysis has been eliminated.

3. *The approach for informing model structure does not appear very elaborate to me. Since this is the core of "model building for understanding catchment process" I particularly miss a clear and elaborate discussion on how the identified landscape characteristics help in the model building process. More specifically:*

   a. *In chapter 3.1.3. you state that the results of the regression analysis were used to build the hydrological model e.g. the subdivision of the catchment in HRUs (p. 7 l. 32). Later, in 4.1.1 you state that subdivisions were defined by gauge locations (p. 11 l 26). I did not find information on how you derived the number of HRUs and the role of catchment characteristics in this context? Chapter 4.1.1. should be more comprehensive in this regard.*

   Our intention was to present chapter 4.1.1 as a general overview of the model structure in order to make the following clearer. The information from the regression analysis are used to derive the HRUs is described in chapters 3.3 and 4.1.5. It is important to make clear the difference between the division in subcatchments (areas that have uniform inputs) and HRUs (areas that have the same hydrological response). The former are defined by the presence of gauging stations (and this division is not negotiable) while the latter reflect our understanding of the catchment functioning (and, in this study, of the regression analysis).

Sections 3.3, 4.1.1, and 4.1.5 in the new version of the paper have been restructured in order to clarify the difference between subcatchments and HRUs and to make the connection between the findings of the correlation analysis and the modeling experiments clearer.

b. *The argument that "the regression analyses have indicated that precipitation is a dominant control on average streamflow" (p. 12 l. 4) is trivial. I don't think you need this and particular not as a justification for using spatially distribution rainfall as a model input. From your manuscript it appears to me that the spatial discretization of your model was based on the definition of the subcatchments (which are in turn defined by the location of gauges) and according the definition of fields (definition not clear). In consequence I don't see that landscape characteristics played an important role in this process. Please clarify?*

   Although it may be a priori clear that precipitation needs to be distributed per subcatchment, it may be not as taken for granted that this is sufficient to capture the water balance of the subcatchments, as many other aspects could in principle play a role (e.g. regional groundwater flow). Here we show that considering distributed precipitation over the subcatchments (defined by the presence of the gauging stations) could by itself be sufficient. Other landscape characteristics play a role in the definition of the HRUs.

   Sections 4.1.1 and 4.1.5 in the new version of the paper have been modified to address this comment; in particular, M0 shows that distributing only the precipitation without accounting for snow related processes is sufficient to capture the average streamflow.

c. *You also mention that "the parameters were motivated by the results of the regression analysis" (p. 8 l. 1). Please omit or explain more detailed. A matrix to illustrate the relationships between model parameters and catchment descriptors would be good. I would for instance be interested in how one could use catchment descriptors to derive (or at least constrain) model storage (kFR or kSR) or network lag (trise,IL trise,OL) parameters. Please comment on that*

   All the process of building the model was motivated by the results of the regression analysis (in particular the decisions on the division in HRUs). The parameters are just calibrated using streamflow data (section 4.1.3). No inference of the parameters from catchment characteristics was done.

   We have changed the misleading sentence (lines 17-20, page 9 of the new version of the paper)

d. *Chapter 4.1.5 is difficult to me due to different reasons: i) your analysis does not VERIFY that "models that account for influencing factors … lead to an improved representation".*

Essentially it only shows that a complicated model (with a larger number of degrees of freedom) outperforms a simpler model (with a smaller amount of degrees of freedom). Please use precise wording. It addresses the question of adequate model complexity. If a lumped representation (M1) is not adequate than also the comparison of M2 to M1 is not adequate. Please explain in more detail why you consider M1 a suitable reference? ii) Please explain why unconsolidated areas receive an individual HRU and why consolidated and alluvial areas can be lumped together (what are your expectation on the underlying processes)? iii) The parameterization of M3 is based on land use, which is not considered to have a causal relationship to the streamflow signatures (Table 2). Please explain why a model which is derived from non-causal properties can be a considered a meaningful reference? Why did you group based on geology and not on elevation, slope or the aridity index which you considered to have a causal relationship? This would maybe be a more appropriate benchmark? iv) Essentially chapter 4.1.5 addresses the questions of optimal degree of model complexity and optimal degree of spatial discretization - which are both very important. However, these aspects are treated together and not separated from each other. Moreover, potential answers to these questions miss a clear link to catchment descriptors. Essentially only differences in geology were considered in the model building. Please clarify to novelties of your study more clearly.

Essentially the two main model configurations are M1 and M2: the first is the baseline and it is a semidistributed model (in the sense that the inputs are spatially distributed and the routing between subcatchments is explicitly addressed in the model) with only one HRU (meaning that all the catchment responds in the same way to the forcings); the second extends the first providing a subdivision of the subcatchments in two HRUs. M3 is used to show that the subdivision in HRUs has to be carefully made otherwise a more complex model doesn't imply automatically better results. Answering to the specific points:

i) M2 is indeed more complex than M1 but our thesis is that its better performance is not just due to the fact that it is more complex but to the fact that it incorporates the right catchment characteristics. This is also demonstrated by M3 that is as complex as M2 but it has the same deficiencies of M1. M1 is already a quite complex model since it already considers the spatial distribution of the inputs and incorporates information about the routing between subcatchments. The real baseline would have been a lumped model, with uniform input and no information about the catchment characteristics but it was too simple for the comparison.

ii) There is an error in the text: the two HRUs are unconsolidated and alluvial (HRU1) vs consolidated (HRU2). Alluvial and unconsolidated geology were put together because they show a similar behavior in terms of water dynamics in the sense that they both represent areas with high storage capacity, especially if compared with HRU2 that is quite impermeable.

iii) M3 was designed to demonstrate that M2 outperforms M1 not just because it is more complex but only because it incorporates characteristics that actually have an

impact on the response of the catchment. For this reason we used a model with the same complexity of M2 but based on characteristics that don't correlate with the streamflow signatures. Also topography was considered in the modeling experiment (but not reported in the paper), experimenting with a subdivision in HRUs based on the slope, but the model resulted similar to M2 (in terms of spatial discretization) but slightly worse in terms of signatures representation. The meteorological characteristics are known at subcatchment scale and therefore, due to the configuration of the model, they are not suitable for the subdivision in HRUs.
We have completely restructured the section 4.1.5 with the intent of making clearer the differences between the modeling experiments and the reasoning behind them. We have also introduced a new model (M0) to test the effect of snow on the seasonality patterns.

e. *the whole structure of the model building story is a bit complicated as aspects are described in chapters 3.1.3, 3.3, 4.1.4 und 4.1.5 which makes it difficult to follow. I suggest combining them into a single chapter. Therein start with the theory e.g. snow is important followed by the surrogate you considered it e.g. half stream flow period. Or geology is important due to… –> different HRUs.*
It was divided in different paragraphs along the paper in order to emphasize the connection between data analysis and modeling choices but we understand that it makes more difficult to follow the story.
Sections 3.1.3, 4.1.1, and 4.1.5 have been restructured in order to improve their readability and to show more clearly the connections between the three sections.

4. *Several conclusions are not appropriate: e.g. "the proposed approach is useful in the fact that modelling can be used to test specific hypotheses on dominant processes resulting from regression analysis" (p. 19 l. 4). This has not be shown. More over aspects related to the event scale are mentioned in the first three bullet points but not subject to the manuscript. In the third bullet point you state: "Higher proportion of consolidated material has an influence on the baseflow vs quickflow portioning, causing lower baseflow and higher peaks" (p 19 l 14). Does the study provide evidence for this statement or does it support this hypothesis? I expect the latter and missed this statement in the chapter 3.1.1. I suggest re-writing of the entire section and to differentiate concisely between hypothesis, results and conclusions.*
With respect to the first point, we think that the model comparisons have been useful to confirm the interpretations of the regression analysis. Clearly the regression between variables is also a model, but the hydrological model is an integrated model that is meant to explain all dependencies at once, whereas the regression model provides a separate model (regression) for each of the dependencies. Therefore there is an added value in the hydrological model, compared to the regression model.
The conclusions of the paper have been revised avoiding aspects related to the event scale and preferring an analysis of the signatures.
5. *The model performance evaluation (chapter 4.1.4) is complicated but of minor importance in this context. I suggest shorting the evaluation section and to focus on a single, interpretable metric*

e.g. the Kling-Gupta Efficiency as the NASH has several limitations and the normalized log-likelihood is difficult to interpret. But this is a minor point and a matter of taste.
We agree that the Nash-Sutcliffe efficiency has several limitations, as any individual index is somehow limited. This is why we have introduced several signatures to evaluate model results. Indeed, we could see that a significant improvement in some of the signatures could result into a negligible improvement in the Nash- Sutcliffe efficiency.

Technical corrections (figures and tables only) I only provide technical corrections for the figures and tables as I expect that several parts of the manuscript will be subject to major revisions.

Thank you for the comments for improving the quality of our figures and tables; we will address them bellow.

- *Figure 1: A: I suggest to remove the colour code and to provide notations (abbreviation) in or around the map. This would help improve the readability of the stream network and the location of the gauges. If you want to keep the legend please add catchment abbreviations to it, order it according to Fig 2. and use a meaningful colour code (e.g. mean annual precipitation, elevation or geology), B: Try a discrete legend like in atlases, will improve readability. C: Forest and pasture are hardly distinguishable both on my screen and in a printed version.*
  Figure 1A: We agree that there are some problems with the readability of the river network but they are mainly due to the poor resolution. In the final version we will upload the figures separately with an higher resolution. The presence of the legend doesn't make the figure smaller since the constraint is the height and not the width of the panel; The colors used for the single catchments were chosen from a "categorical" color scale in order to be as different as possible. Linking them to some characteristic would mean using a "sequential" color scale, with little difference between subcatchments, and this would be problematic in the other figures (assuming that we want to be consistent) where we want to clearly see the behavior of the single catchments.
  Figure 1A : we have changed the figure keeping only the main rivers in order to improve the readability.
  Figure 1B: we have improved the figure according to the suggestion
  Figure 1C: we have changed the colors (darker green for the forest) to improve the readability.
- *Figure 2: Please repeat the variables and their abbreviation in the caption such that the figure can be read independent from the text. Maybe add another row and provide grouping indices based on the results in chapter 3.2.1*
  As the names are relatively long, they would not fit on the y axes. Instead, we have opted to place them in the title of the subplots.
  Figure 2 is not present in the new version of the paper.
- *Figure 3: Please repeat the variables and their abbreviation in the caption such that the figure can be read independent from the text.*
  See reply at earlier point.
  Figure 3 is not present in the new version of the paper.

- *Figure 4: Please repeat the variables and their abbreviation in the caption such that the figure can be read independent from the text. Information on the range of the different variables would be pretty helpful as well. If possible include it otherwise please mention the ranges in the text or add the information to table 1.*
  The range of the variables is always between 0 and 1: all the variables plotted are percentage of the area of the subcatchment occupied by a certain characteristic. The characteristics that don't belong to the category "part of the catchment occupied by …" are reported in table 1.
  Figure 4 is not present in the new version of the paper.
- *Figure 5: I'm not sure if this figure is required since B and C show very little variation. The only important message from A is that there a catchments that are stronger controlled by snow than others. I suggest removing it. If you decide to keep it update the colour code according to the suggestion for Figure 1.*
  Although the plots B and C show little variability across the catchment, it is still interesting to present the seasonal dynamics. Moreover, we consider that it is useful to show that the monthly variability in streamflow (plot A) is not directly ascribed to variability in precipitation or potential evaporation (plot B or C).
  Figure 5 is not present in the new version of the paper.
- *Figure 6: I cannot find a description of the symbols and abbreviations in the Appendix. Please specify at least the meaning of the capital letters in the caption (as in 4.1.1) and provide a more comprehensive description in the appendix.*
  We have put the description of the abbreviations in the caption of the figure
- *Figure 7: Order according to Figure 2 or 3. Line type and colour code are redundant.*
  We have added M0 to figure 7. The order of the catchment has been kept alphabetical since there is no more need to be consistent with figures 2 and 3.
- *Figure 8, 9, 10: Nice figures! Suggestions: Combine all three figures in one (each model setup as an individual row). This would improve readability. Streamflow, runoff coefficient and half streamflow period have no or little variation (two out of these could be omitted such that all results would fit in one figure). Remove the correlation coefficients due to their distracting nature (correlation (alone) is pretty meaningless in this context). Update colour code according to Fig 1 A.*
  Point taken. We acknowledge that is more meaningful to put the different models together in order to facilitate the comparison.
  We have put all the models side by side in the new figures 9 and 10.
- *Table 1: Order according to fig 1 A. Index column is not relevant, omit Code or put it to the very right. Rounding is not yet meaningfully and consistent.*
  The Index column is used in the "upstream catchments" column to define the river network. The "code" column is present to avoid ambiguity with the naming of the gauging stations providing the reader with the code of the gauging station used by the Federal Office for the Environment FOEN.

The new version of Table 1 has a reduced number of columns and its primary goal has changed from describing the catchment characteristics to identifying the gauging stations and to describe the river network.

- *Table 2: This table includes variables with spurious correlations (Brett 2004, Kenny 1982). This includes variables that are considered statistically significant and where causality was assumed e.g. the correlation between aridity index AI and the runoff coefficient RC which are both are derived from precipitation. The same applies for P and RC. Since P and Q are highly correlated and AI is based on P I also wonder about the significance of AI and RC, BFI, FI and HDP. Please clarify. Please also explain why you assumed causality among LP and BFI and among LP and FI? Differences among the geological fractions are small as well. Why do you consider causality in some of the individual relationships and in others not?*
  Table 2 is not present in the new version of the paper.
- *Table 3: This analysis also includes variables with spurious correlations. Please comment on that.*
  Table 2 is not present in the new version of the paper.
- *Table A1: Please provide a brief explanation on parameters and components. Where does the range of variability come from?*
  Table A1 has been modified to address this comment.
- *Table A2: Explain component*
  Table A2 has been modified to address this comment.

Literature

Brett, M. T. (2004). When is a correlation between non-independent variables "spurious"?

Oikos, 105(3), 647–656. Kenney, B. C. (1982). Beware of spurious self-correlations! Water Resources Research, 18(4), 1041–1048. https://doi.org/10.1029/WR018i004p01041

Reply to review by Lieke Melsen

We thank Dr. Lieke Melsen for her careful read to the manuscript and insightful suggestions. Below, we answer in detail the various comments, and illustrate how we have addressed them in the revised version. The original comments of the reviewer are reported in *black and italics*, our replies in blue and the changes to the paper in **green and bold.**

Note that, were not indicated, the numbers of figure/tables/lines/pages refer to the old version of the paper.

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. 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.
   The list of signatures and indices has been extended in the new version of section 3.1.1. taking advantage of previous works.

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.

Section 3.2.2 in the new version of the paper deals with mutual correlations with the objective of selecting signatures and indices that are either not correlated or that represent different characteristics.

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 form 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.

   The number of signatures and indices has been reduced using the "internal" correlation analysis and expert judgment (section 3.1.2 with results in 3.2.1 of the new version of the paper). This process has simplified the selection of influencing factors on streamflow signatures (section 3.1.3 with results in 3.1.2 of the new version of the paper). In both cases, we have highlighted the role of "expert judgment" in the process of indices selection.

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.

   In the new version of the paper all the analyses have been based on Spearman correlation.

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.

The explanatory power of the signatures has been highlighted (section 3.1.1 of the new version of the paper) and, in particular, we have made the hypothesis about the importance of the snow more explicit introducing the model M0.

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 where 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 have made clearer what we mean for model building, especially restructuring sections 4.1.1 and 4.1.5.

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.

Section 3.2.1 has been completely revised in the new version 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 has been completely revised in the new version of the paper.

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.

We have tried in the new version of the paper to balance the usage of symbols with the usage of the full name.

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).

Several figures have been changed in the new version of the paper.

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 have clarified that (section 3.1.1 in the new version of the paper)

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.

The actual meaning is "optimizing the parameters of the hydriological 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"

We have changed the sentence (line 1, page 14 of the new version of the paper)

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 have done 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.

Data analysis and model building for understanding catchment processes: the case study of the Thur catchment.

Marco Dal Molin[1,2,3], Mario Schirmer[2,3], Massimiliano Zappa[4], Fabrizio Fenicia[1]

[1]Department Systems Analysis, Integrated Assessment and Modelling, Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland
[2]The Centre of Hydrogeology and Geothermics (CHYN), University of Neuchâtel, 2000 Neuchâtel, Switzerland
[3]Department of Water Resources and Drinking Water, Eawag, Swiss Federal Institute of Aquatic Science and Technology, 8600 Dübendorf, Switzerland
[4]Hydrological Forecast, Swiss Federal Research Institute WSL, 8903 Birmensdorf, Switzerland

Correspondence to: Marco Dal Molin (marco.dalmolin@eawag.ch)

Abstract

The development of semidistributed hydrological models that reflect the dominant processes controlling streamflow spatial variability is a challenging task. This study ~~addresses~~illustrates this ~~problem by investigating~~process through the case of the Thur catchment (Switzerland, 1702 km^2), an alpine and pre–alpine catchment ~~that, while having a moderate (1702 km^2) extension, presents a~~with large spatial variability in terms of climate, landscape, and streamflow (measured at 10 subcatchments). ~~The methodology for~~In order to appraise the dominant processes that control catchment response, and build a model that reflects them, the model development ~~consists of~~follows a two–stages approach. In a first stage, we use correlation ~~and regression~~ analysis to identify the main influencing factors on the spatial variability of streamflow signatures. Results of this analysis show that precipitation ~~(rainfall or snow) controls~~averages control signatures of ~~seasonality and~~ water balance, snow processes control signatures of seasonality, while landscape characteristics (especially geology) control signatures ~~of hydrograph shape (e.g.~~ characterizing the importance of baseflow ~~index and flashiness index).~~. In a second stage, ~~we use~~ the results of the previous analysis are used to develop a ~~semidistributed hydrological model that is consistent with the data. Model~~ set of model experiments aimed at determining an appropriate model representation of the Thur catchment. These experiments confirm that only a hydrological ~~models~~model that ~~account~~accounts for the heterogeneity of precipitation ~~and~~, snow related processes, and landscape features such as geology ~~produce~~, produces hydrographs that have signatures similar to the observed ones. ~~These models provide~~This model provides consistent results in space–time validation, which is promising for ~~prediction~~predictions in ungauged ~~conditions~~basins. The presented methodology for model building can be transferred to other case studies, since the data used in this work (meteorological variables, streamflow, morphology and geology maps) is available in ~~many~~numerous regions around the globe.

1 Introduction

~~Hydrographs are affected by meteorological forcing and landscape characteristics (e.g. topography, land use, etc.) and, therefore, they synthetize the hydrological response of a catchment. Because of~~Due to the spatial variability of landscape (e.g. topography, land use, etc.) and climate ~~characteristics~~, hydrographs can differ substantially between catchments. Being able to quantify and explain hydrograph spatial variability is important both to improve processes understanding and to make predictions useful for many human activities, such as flood protection, drinking water production, agriculture, energy production, and riverine ecosystems management (e.g., Hurford and Harou, 2014).

Understanding catchment differences and, more specifically, how to transfer hydrological knowledge, methods, and theories ~~from one place to another~~between places, is a common objective of many research areas in hydrology, including comparative hydrology (e.g., Falkenmark and Chapman, 1989), model regionalization (e.g., Parajka et al., 2005), catchment classification (e.g., Wagener et al., 2007), and prediction in ungauged basins (e.g., Hrachowitz et al., 2013). In the case of streamflow, the attempt to explain its spatial variability is typically accomplished either using statistical approaches, which ~~try~~are designed to regionalize ~~some~~selected characteristics of the hydrograph (streamflow signatures), or ~~using~~through hydrological models that ~~incorporate~~account for relevant spatial information. In particular, statistical approaches such as regression analysis (e.g., Berger and Entekhabi, 2001; Bloomfield et al., 2009) and correlation analysis (e.g., Trancoso et al., 2017), or machine learning techniques like clustering (e.g., Sawicz et al., 2011; Toth, 2013; Kuentz et al., 2017) are used to ~~extrapolate the signatures where unknown and to~~ group together catchments that present similar characteristics and to extrapolate the signatures where unknown. Such approaches have been useful to quantify the hydrological variability and identify its principal ~~divers~~drivers. However, they are often not designed to discover causality links and can be affected by multicollinearity, that arises when multiple factors are correlated internally and with the target variable (Kroll and Song, 2013).

By incorporating spatial information about meteorological forcing and landscape characteristics, ~~distributed~~semidistributed hydrological models have the ability to ~~reproduce~~mimic the mechanisms that influence hydrograph spatial variability. However, identifying the relevant mechanisms is challenging. One possibility is to be as inclusive as possible in accounting for all the catchment properties that are, in principle, important in controlling catchment response. However, this approach leads to models that tend to be data demanding and contain ~~several~~many parameters. For example, Gurtz et al. (1999) considered several landscape characteristics (elevation, land use, etc.) in their application of a semidistributed model to the Thur catchment (Switzerland), which resulted into a model with hundreds of hydrological response units (HRUs) that were defined a–priori based on the complexity of the catchment. The other option is to try to identify the most relevant processes and neglect others, ~~by tuning the distributed hydrological model to the available data.~~in order to control model complexity. For example, Fenicia et al. (2016) compared various model hypotheses to determine an appropriate discretization of the catchment in HRUs and appropriate structures for different HRUs. ~~However, in their work, the space of plausible hypotheses could be constrained by a good experimental understanding of the area, which is not always available.~~Antonetti et al. (2016)

used a map of dominant runoff processes following Scherrer and Naef (2003) for defining HRUs. However, these approaches require a good experimental understanding of the area, which is not always available.

Convincing model calibration–validation strategies are essential to provide confidence that the model ability to fit observations is a reflection of model realism and not a consequence of calibrating an overparameterized model (e.g., Andréassian et al., 2009). A common approach for calibration of semidistributed models is the so called 'sequential' approach, where subcatchments are calibrated sequentially from upstream to downstream (e.g., Verbunt et al., 2006; Feyen et al., 2008; Lerat et al., 2012; De Lavenne et al., 2016). Although this approach may provide good fits and therefore it has its practical utility where data is available, it does not provide understanding into the causes of streamflow spatial variability and results into models that are not spatially transferable. Moreover, such models are prone to contain many parameters, as each subcatchment would be represented by its own set of parameters ~~set~~. Alternative calibration–validation approaches that enable model validation not only in time but also in space are conceptually preferable, particularly when the modeling is used for process understanding or prediction in ungauged locations (e.g., Wagener et al., 2004; Fenicia et al., 2016).

This study combines the strengths of catchment regionalization approaches and ~~distributed~~semidistributed hydrological models by first using ~~regionalization studies~~regression analysis to understand the main causes of variability of streamflow signatures, and then using this analysis to inform the structure of a distributed hydrological model. The model objective is to explain the observed spatial diversity of streamflow characteristics with the minimum possible complexity, while maintaining a process based interpretation. In particular, the objectives of the study are to: (1) explore the spatial variability present in the Swiss Thur catchment regarding landscape characteristics, meteorological forcing and streamflow signatures; (2) ~~find which characteristics~~identify the main drivers that explain the variability of the hydrological response; (3) based on this analysis, build a ~~semidistributed hydrological~~set of model ~~that considers only features that actually contribute~~experiments aimed to ~~the spatial variability~~test the relative importance of dominant processes and their effect on the hydrograph; (4) ~~validate~~appraise model assumptions against competing alternatives using a stringent validation strategy.

The paper is organized as follows: Section 2 presents the study area and gives information about data ~~collection and~~ availability; Section 3 and Sect. 4 are both divided in methods and results and present, respectively, the correlation ~~and regression~~ analysis and the modeling part of this paper; Section 5 puts the results of this work in ~~prospective~~perspective, comparing them with other studies; Section 6, finally, summarizes the main conclusions.

2 Study area

This study is carried out in the Thur catchment (Fig. 1), located in north–east of Switzerland, south–west of the Lake Constance. With a total length of 127 km and a catchment area of 1702 km^2, the Thur is the longest Swiss river without any natural or artificial reservoir along its course. ~~Due to this characteristic, it~~The Thur river is ~~a~~ very dynamic ~~river, where the~~, with streamflow values that can change ~~of~~by two orders of magnitude ~~in~~within a few hours (Schirmer et al., 2014).

The Thur catchment has been subject of several studies in the past; Gurtz et al. (1999) ~~did~~performed the first modelling study on the entire catchment using a ~~semidistributed~~semi-distributed hydrological model; Abbaspour et al. (2007) modelled hydrology and water quality using the SWAT model; Fundel et al. (2013) and Jorg-Hess et al. (2015) focused on low flows and droughts; Jasper et al. (2004) investigated the impact of climate change on the natural water budget. Other modelling studies include also Melsen et al. (2014) and Melsen et al. (2016), which investigated parameters estimation in data limited scenarios and their transferability across spatial and temporal scales, and Brunner et al. (2019) ~~that~~who studied the spatial dependence of floods. The Thur includes also a small–size experimental subcatchment (Rietholzbach, called Mosnang in this paper after the name of the gauging station) that was subject of many field studies at the interface between process understanding and hydrological modelling (e.g., Menzel, 1996; Gurtz et al., 2003; Seneviratne et al., 2012; von Freyberg et al., 2014; von Freyberg et al., 2015).

The topography of the catchment is presented in Fig. 1b; the elevation ranges between 356 m a.s.l. at the outlet and 2502 m a.s.l. at Mount Säntis. The majority of the catchment lies below 1000 m a.s.l (75 %) and only 0.6 % is above 2000 m a.s.l. (Gurtz et al., 1999). Based on topography (Fig. 1b~~,~~), the catchment can be visually subdivided into two distinct regions: the northern part, with low elevation and dominated by hills and flat land, and the southern part, ~~that~~which presents a mountainous landscape. ~~Such topographic variability suggests the presence of different dynamics in precipitation (type and quantity) and routing.~~

The land use (Fig. 1c) is dominated by pasture and sparse vegetated soil (60 %) and forest (25 %); urbanized and cultivated areas are located mainly in the north and cover~~, respectively, the~~ 7 % and ~~the~~ 4 % of the catchment respectively.

Most of the catchment is underlain by conglomerates, marl incrustations and sandstone (Gurtz et al., 1999). For the purpose of this study, the geological formations have been divided into ~~3~~three classes (Fig. 1d): “consolidated”, covering mainly the mountainous part of the catchment, “unconsolidated”, located in the north, and “alluvial”, located in the proximity of the river network, mainly in the plateau area; the latter formation constitutes the main source of groundwater in the region (Schirmer et al., 2014). The soil depth (Fig. 1e) is shallower in the mountainous part of the catchment and deeper in the northern part.

Based on the availability of gauging stations~~,~~ (Table 1), the catchment was divided in 10 subcatchments (Fig. 1a), with a total drained area that ranges between 3.2 km^2 (Mosnang) and 1702 km^2 (~~total catchment area).~~Andelfingen). Streamflow time series are obtained from the Federal Office for the Environment FOEN and the data is available from 1974 to 2017 but is used only form 1981 to 2005 to match the precipitation, temperature, and potential evapotranspiration (PET) time series. In the considered range, the streamflow data are relatively continuous, with two gaps, one in St. Gallen, from 31 December 1981 to 01 January 1983, and the other one in Herisau, from 31 December 1982 to 09 May 1983.

The raw maps (topography, land use, geology, and soil) are obtained from the Federal Office of Topography swisstopo. The meteorological data is obtained from the Federal Office of Meteorology and Climatology MeteoSwiss. Precipitation and temperature are interpolated, as done in Melsen et al. (2016), with the pre–processing tool WINMET (Viviroli et al., 2009) using~~, respectively,~~ inverse distance weight (IDW) and detrended IDW respectively; while the first method considers only the horizontal variability (related to the distance from the meteorological stations), the second adds a vertical component to the variability related with the elevation (Garen and Marks, 2001). PET data is then obtained, as done in Gurtz et al. (1999), starting from meteorological and land use data, using the Penman–Monteith equation (Monteith, 1975), implemented as part of the hydrological model PREVAH (Viviroli et al., 2009). All these values are calculated at pixel (100 m) scale and then averaged over the subcatchments. ~~All the time series are used at daily time step, aggregating the available hourly data.~~All the time series are used at daily resolution in the subsequent analyses, aggregating the available hourly data. This choice was influenced on one hand by the need of limiting the computational demand for the model experiments, for which a coarser temporal resolution is preferable, and on the other hand by the need of representing relevant hydrograph dynamics, for which finer temporal resolution is desirable (e.g., Kavetski et al., 2011). A daily data resolution, although it may obscure subdaily process dynamics, appeared to be a good compromise, and it is a typical choice in distributed model applications at such spatial scales (e.g., Kirchner et al., 2004).

~~The raw maps (topography, land use, geology, and soil) are obtained from the Federal Office of Topography swisstopo. Visualization and processing is performed using Qgis 2.18.~~

3 Identification of influencing factors on the spatial variability of streamflow signatures

3.1 Methodology

The purpose of the analysis presented in this section is to understand the influence of climatic conditions and landscape characteristics on streamflow. Climatic conditions are represented by precipitation~~ and~~, potential evaporation~~ data.~~, and temperature time series. Landscape characteristics are ~~presented~~represented by maps of topography, land use, geology and soil.

Climatic conditions, landscape characteristics and streamflow are represented through a set of ~~indices, designed with the intention of being representative of the underlying data.~~statistics. In the following, ~~indices~~statistics calculated based on streamflow data will be called streamflow “signatures”, as it is often done in catchment classification literature (e.g., Sivapalan, 2006). ~~Dependencies between streamflow signatures and other indices are assessed through a regression analysis. The signatures used are illustrated in Sect. 3.1.1, the regression analysis between streamflow signatures and the other indices is explained in Sect. 3.1.2, and the guidelines for interpretation are given in Sect. 3.1.3~~We will refer to climatic and landscape indices for statistics calculated on climatic data and landscape characteristics. A broad list of signatures and indices is presented in Sect. 3.1.1; Section 3.1.2 presents an approach for reducing such list to a set of meaningful variables; Section 3.1.3 illustrates the approach for determining the indices that mostly control streamflow signatures.

3.1.1 Catchment indices for representing streamflow, climate, and landscape

Streamflow signatures (ζ) and ~~meteorological~~climatic indices (ψ) were obtained using streamflow, precipitation, PET, and ~~PET~~temperature time series~~ at daily time step.~~. The values were calculated ~~for each year, starting on~~using 24 years of data, between 01 September~~, here~~ 1981 and 31 August 2005; we considered the 01 September as the beginning of the hydrological year~~, and then averaged over the entire period; years~~. The periods with gaps in the data (refer to Sect. 2 for details) were ~~completely~~ discarded from the analysis of the specific subcatchment. Landscape indices were obtained using the maps described in Section 2.

~~Streamflow was represented through the following signatures:~~

Addor et al. (2017) recently compiled a comprehensive list of streamflow signatures and climatic indices for characterizing catchment behaviour (see Table 3 in Addor et al. (2017)). Here, we adopted their selection. The streamflow signatures here considered are described hereafter, followed by an explanation of their rationale:

- average daily streamflow ~~(ζ_Q~~ $\zeta_{\mathrm{Q}} = \bar{\boldsymbol{q}}$~~),~~, where $\boldsymbol{q}$ is the streamflow time series and the overbar represents the average over the observation period;
- runoff ~~coefficient~~ratio ~~ζ_{RC}~~ $\zeta_{\mathrm{RR}} = \frac{\bar{\boldsymbol{q}}}{\bar{\boldsymbol{p}}}$, where $\boldsymbol{p}$ is the precipitation time series;
- ~~baseflow index $\zeta_{BFI} = \frac{\overline{\boldsymbol{q}^{(\boldsymbol{b})}}}{\bar{\boldsymbol{q}}}$, where $\boldsymbol{q}^{(\boldsymbol{b})}$ represents the baseflow and was calculated using a low pass filter as illustrated in (Eckhardt, 2008), Eq. (5)~~

  $$\cancel{q_t^{(b)} = \min\left(q_t, \vartheta_B q_{t-1}^{(b)} + \frac{1-\vartheta_B}{2}(q_{t-1} + q_t)\right).} \quad \cancel{(1)}$$

  ~~According to Eckhardt (2008), a single forward filter pass was applied but the parameter ϑ_B was chosen to be equal to 0.99, instead of 0.925 (suggested by Eckhardt, 2008), to highlight the low frequency component of the hydrograph. It was found that the choice of ϑ_B, although affecting the baseflow index (BFI) of individual subcatchments, it did not change their relative values significantly. This choice therefore had a limited influence on the results of the regression analysis;~~
- ~~flashiness index (Baker et al., 2004), defined as~~

  $$\cancel{\zeta_{FI} = \frac{\sum_{t=2}^{N_T}|q_t - q_{t-1}|}{\sum_{t=2}^{N_T} q_t};} \quad \cancel{(2)}$$

  ~~and used to describe the "responsiveness" of a catchment.~~
- streamflow elasticity (ζ_{EL}) defined as

  $$\zeta_{\mathrm{EL}} = \mathrm{med}\left(\left(\frac{\Delta\bar{\boldsymbol{q}}}{\bar{\boldsymbol{q}}}\right)/\left(\frac{\Delta\bar{\boldsymbol{p}}}{\bar{\boldsymbol{p}}}\right)\right) \quad (1)$$

  where $\Delta\bar{\boldsymbol{q}}$ and $\Delta\bar{\boldsymbol{p}}$ represent the streamflow and precipitation jumps between two consecutive years and med is the median function;
- slope of the flow duration curve (ζ_{FDC}) defined as the slope between the log-transformed 33rd and 66th streamflow percentiles;
- baseflow index $\zeta_{\mathrm{BFI}} = \frac{\overline{\boldsymbol{q}^{(\mathbf{b})}}}{\bar{\boldsymbol{q}}}$, where $\boldsymbol{q}^{(\mathbf{b})}$ represents the baseflow and was calculated using a low–pass filter as illustrated in Ladson et al. (2013) with the equation

$$q_t^{(f)} = \min\left(0, \vartheta_b q_{t-1}^{(f)} + \frac{1+\vartheta_b}{2}(q_t - q_{t-1})\right) \quad (2)$$

$$q_t^{(b)} = q_t - q_t^{(f)} \quad (3)$$

with $q_t^{(f)}$ representing the quick flow. The settings of the filter were taken according to the findings of Ladson et al. (2013) and, in particular, three filter passes were applied (forward, backward, and forward), the parameter ϑ_b was chosen to be equal to 0.925, and a reflection of 30 time steps at the beginning and at the end of the time series was used;

- mean half streamflow ~~period~~date ($\cancel{\zeta_{HSP}}\zeta_{\mathrm{HFD}}$) (Court, 1962), defined as the number of days needed in order to have a cumulated streamflow that reaches the 50 % of the total annual streamflow;~~ the value obtained is then normalized by the total number of the days in the year. This index is designed to capture the seasonality of streamflow, since it helps differentiating between catchments with high streamflow during the winter and catchments with high streamflow during the spring.~~

~~Climatology was represented through the following indices:~~

- 5th and 95th percentiles of the streamflow (ζ_{Q5} and ζ_{Q95} respectively);
- frequency (ζ_{HQF}) and mean duration (ζ_{HQD}) of high-flow events: they are defined as the days when the streamflow is bigger than nine times the median daily streamflow;
- frequency (ζ_{LQF}) and mean duration (ζ_{LQD}) of low-flow events: they are defined as the days when the streamflow is smaller than 0.2 times the mean daily streamflow;

The frequency of days with zero streamflow, present in Addor et al. (2017), was not considered in this study because there are no ephemeral subcatchments in the study area.

This group of streamflow signatures is capable of capturing various characteristics of the hydrograph: ζ_{Q} measures the overall water flows, ζ_{RR} represents the proportion of precipitation that becomes streamflow, ζ_{EL} measures the sensitivity of the streamflow to precipitation variations, with a value greater than 1 indicating an elastic subcatchment (i.e. sensitive to change of precipitation) (Sawicz et al., 2011), ζ_{FDC} measures the variability of the hydrograph with a steeper flow duration curve indicating a more variable streamflow, ζ_{BFI} measures the magnitude of the baseflow component of the hydrograph, and can be considered as a proxy for the relative amount of groundwater flow in the hydrograph, ζ_{HFD} measures the streamflow seasonality, ζ_{Q5}, ζ_{LQF}, and ζ_{LQD} measure low-flow dynamics, ζ_{Q95}, ζ_{HQF}, and ζ_{HQD} measure high-flow dynamics.

Climatology was represented through the following indices (see Addor et al. (2017), Table 2):

- average precipitation $\cancel{\psi_P}\psi_{\mathrm{P}} = \overline{\boldsymbol{p}}$;
- ~~average PET~~ $\cancel{\psi_{PET} = \overline{\boldsymbol{e}_{pot}}}$;
- average PET $\psi_{\mathrm{PET}} = \overline{\boldsymbol{e}_{\mathrm{pot}}}$, where $\boldsymbol{e}_{\mathrm{pot}}$ is the potential evapotranspiration time series;
- aridity index $\cancel{\psi_{AI} = \frac{\overline{\boldsymbol{e}_{pot}}}{\overline{\boldsymbol{p}}}}$ $\psi_{\mathrm{AI}} = \frac{\overline{\boldsymbol{e}_{\mathrm{pot}}}}{\overline{\boldsymbol{p}}}$;

- ~~These indices were designed to capture different features~~fraction of snow (ψ_{FS}), defined as the ~~time series: yearly streamflow,~~ volumetric fraction of precipitation falling as snow (i.e. on days colder than 0 °C);
- frequency (ψ_{HPF}) and ~~PET can be called "magnitude" indices since~~mean duration (ψ_{HPD}) of high precipitation events: they are ~~a measure of~~ defined as days when the ~~water flows;~~precipitation is bigger than five times the ~~remaining indices give information about~~mean daily precipitation;
- season (ψ_{HPS}) with most high precipitation events (defined as above);
- frequency (ψ_{LPF}) and mean duration (ψ_{LPD}) of dry days: they defined as days when the ~~"shape" of~~ precipitation is lower than 1 mm day$^{-1;}$
- season (ψ_{LPS}) with most dry days (defined as above).

The seasonality of precipitation used in Addor et al. (2017) was not considered in this study as it relied on fitting a sinusoidal function to the ~~time series~~precipitation values, which in our case did not produce reliable results. Nevertheless, these climatological indices are able to comprehensively represent the climatic conditions of the suubcatchment, with ψ_{P} representing average water input, ψ_{PET} representing average evaporative demand, ψ_{AI} measuring the dryness of the climate, ψ_{FS} measuring the relative importance of snow, ψ_{HPF}, ψ_{HPD}, and ψ_{HPS} measuring the importance of intense precipitation events, and ψ_{LPF}, ψ_{LPD}, and ψ_{LPS} measuring the importance of dry days.

The landscape characteristics~~, illustrated in Sect. 2, need to be synthetized in a numeric value~~ were divided in four categories: topography, land use, soil, and geology. In order to quantify the characteristics of each category, a set of indices (ξ) ~~before being used in the correlation and regression analysis. The maps were processed using GIS techniques and, for each subcatchment, numerical features were extracted. All~~was defined. It is important to notice that all the areas calculated in this analysis were normalized by ~~their~~the respective subcatchment area ($\xi_{\cancel{A}}\xi_{\mathrm{A}}$) in order to get comparable values between subcatchments of different size.

~~In particular,~~Topography was represented with the following indices, calculated based on the digital elevation model (DEM~~)~~ ~~was used to calculate the following topographic information:~~):

- average elevation ($\xi_{\cancel{TE}\mathrm{TE}}$);
- average slope ($\xi_{\cancel{TSm}\mathrm{TSm}}$);
- fraction of the subcatchment with steep areas ($\cancel{\xi_{TSs}}\xi_{\mathrm{TSs}}$) , with slope larger than 10°;
- aspect, i.e. ~~areas~~fraction of the subcatchment facing north ($\cancel{\xi_{TAn}}\xi_{\mathrm{TAn}}$), south ($\cancel{\xi_{TAs}}$~~)~~$\xi_{\mathrm{TAs}}$), or east and west ($\xi_{\cancel{TAew}\mathrm{TAew}}$).

~~The~~Land use was represented with the following characteristics, obtained by reclassifying the land use map ~~was reclassified~~ in four categories (from 22 original classes):

- ~~crops (ξ_{LC});~~
- fraction of the subcatchment with crops land use (ξ_{LC});
- fraction of the subcatchment with pasture land use ($\xi_{\cancel{LP}\mathrm{LP}}$);

- ~~forest (ξ_{LF});~~
- fraction of the subcatchment with forest land use (ξ_{LF});
- fraction of the subcatchment with urbanized land use ($\xi_{\cancel{LU}\mathrm{LU}}$).

~~The soil map was used to quantify:~~

~~areas~~Soil type was represented with the following indices, derived by the soil map:

- fraction of the subcatchment with deep soil (soil depth greater than two meters) ($\xi_{\cancel{SD}\mathrm{SD}}$);
- average soil depth ($\xi_{\cancel{SM}\mathrm{SM}}$).

~~The geology map was reclassified~~Geology was represented by the following indices, obtained by reclassifying the original map in three categories (from 22 original classes):

- fraction of the subcatchment with alluvial geology ($\xi_{\cancel{GA}\mathrm{GA}}$);
- fraction of the subcatchment with consolidated geology ($\xi_{\cancel{GC}\mathrm{GC}}$);
- fraction of the subcatchment with unconsolidated geology ($\cancel{\xi_{GU}}\xi_{\mathrm{GU}}$).

The reclassification of the land use and of the geology maps consisted in aggregating specific classes into general classes (e.g. combining different types of forests into a unique forest class) with the objective of reducing ~~the~~their number~~ of classes~~, in order to facilitate subsequent analyses.

~~**3.1.2 Correlation and regression analysis**~~

~~This analysis is aimed at identifying meteorological and landscape characteristics (ψ and ξ) that mostly control streamflow signatures (ζ). The analysis is subdivided in different steps.~~

~~The first step is about checking whether the Pearson correlation is an appropriate metric for representing correlation. For this purpose, we calculated the correlation between variables using the Pearson correlation coefficient and the Spearman's rank score. The first captures linear correlation while the second is capable of measuring also non-linear (but still monotonic) correlation between two variables (e.g., Artusi et al., 2002).~~

~~The second step in the correlation analysis is aimed at excluding non-significant influencing factors. This selection is based on the following two criteria: (i) the correlations have to be statistically significant, with p-value lower than 0.05; and (ii) the landscape characteristics (ξ) have to cover at least 5 % of the subcatchment. The latter point is~~The consideration of topography, land use, soil, and geology for defining landscape indices was based on their potential influence on hydrological processes, and in turn, on the shape of the hydrograph. These landscape characteristics can all play an important role in controlling hydrological processes: land use can, for example, influence the infiltration of water in the substrate; soil thickness can affect the partitioning between water storage and runoff; vegetation is typically assumed to affect evaporation, and geology can affect groundwater dynamics. Indeed, these characteristics are used by many semidistributed hydrological models, for example for determining parameter values or for dividing the catchment in areas with homogenous hydrological response(e.g., Gurtz et al., 1999).

3.1.2 Selection of meaningful streamflow signatures, climatic indices, and catchment indices

The sets of statistics presented in Sect. 3.1.1 were designed to be comprehensive. However, they may also be redundant, for example by containing metrics that express similar characteristics of the underlying data. In order to facilitate subsequent correlation analyses between the various sets of statistics, it is important to reduce each set to a short list of meaningful variables. The reduction of each set of streamflow signatures, climatic indices, and landscape indices was achieved through the following steps:

- All the statistics that did not show sufficient variability between the subcatchments were eliminated. We were in fact interested in identifying causes of spatial variability in the streafmow dynamics of the subcatchments of the Thur. Therefore, statistics that had a low variability were not of interest in this analysis. The variability was measured using the coefficient of variation (defined by the ratio between the standard deviation and the average) and statistics with a coefficient of variation less than 5 % were discarded.
- All the catchment indices (e.g. a certain type of land use) that account for a limited part of the subcatchment were discarded. The latter point was motivated by the expectation that landscape characteristics covering a very small fraction of the subcatchment should not have a strong influence on the streamflow signatures ~~here considered. Characteristics at point (ii) could already have been excluded before the analysis. Nevertheless we thought it would still be interesting to see the complete picture of correlation between variables~~considered. Here, landscape indices accounting for less than 5 % of the subcatchment area were discarded.

~~The third step in the correlation analysis aims at distinguishing causality from mere correlation. The identification of causality links is based on expert judgment. For example, average streamflow may have a high correlation both with average altitude and with average precipitation, but if high mountain regions also have higher precipitations due to orographic effects, the cause of spatial streamflow variability will be precipitation and not altitude. Additionally, if more indices representing the same landscape characteristic (e.g.~~ ξ_{GA}, ξ_{GC}, ~~and~~ ξ_{GU} ~~are complementary representations of the geology) are correlated with a signature, then only the one with the highest correlation is considered.~~

~~As a last step, in order to determine a range of influencing factors on each of the signatures, we used a linear regression analysis with forward selection of the variables (Miller et al., 2002). In particular, for each signature, only the indices~~ ψ ~~and~~ ξ ~~that exhibit causality were used in a decreasing order of correlation. Starting from the null hypothesis (none of the characteristics is necessary to explain the signatures), at each step a variable was added in the regression and the change in the performance was assessed evaluating the variation of the squared correlation (~~r^2~~) and the residual sum of squares (RSS). The change between these metrics in the steps of the regression was used to interpret the explanatory power of the added variable.~~

~~**3.1.3 Approach for informing model structure**~~

- ~~The results of the regression analysis were used to build the hydrological model; its definition~~Within each set of streamflow signatures, climatic indices, and catchment indices we retained only relatively independent metrics. This step was motivated by the need of removing redundant information within each set. The selection of independent metrics was aided by the Spearman's rank score between each pair of metrics, which represents (also non–linear) correlation between variables. Pairs of metrics with high Spearman's rank score are potentially redundant. In eliminating potentially redundant variables, we adopted the following criteria:
    - Among highly correlated metrics, we preferred those depending on single variables (e.g. only precipitation or only streamflow) to those containing multiple variables (e.g. combining precipitation and streamflow or evaporation, such as the aridity index or the runoff ratio), as this may be a problem when looking for correlations between metrics;
    - With respect to landscape indices, in many cases the high correlation is due to the fact that they are complementary (the areal fractions sum up to unity). In such cases, we kept one index per class (e.g. a single index for geology).
    - A high correlation between metrics does not always mean that the metrics represent the same information. Therefore, the final selection of relevant metrics within each set was guided by expert judgment.

Based on this process, we compiled a reduced list of signatures, climatic indices, and landscape indices, which was used in subsequent analyses.

3.1.3 Identification of climate and landscape controls on streamflow and consequences for model development

This analysis aimed to identify climatic and landscape indices that mostly control streamflow signatures. In order to identify causality links between indices (ψ and ξ) and signatures (ζ) we proceed as follows:

- We calculated the correlation between indices and signatures using the Spearman's rank score, and identified pairs of variables with high correlation;
- We scrutinized pairs of variables with high correlations using expert judgment to decide if a causality link between variables is justified;
- We used the identified causality links to inform the structure of a distributed model.

The distributed model development involved a series of choices regarding the subdivision of the catchment in HRUs, the model structure, and the parameters~~ that~~: all these choices were, in this study, ~~were~~ motivated by the results of the ~~regression~~correlation analysis, i.e. only catchment characteristics that were found capable of explaining the hydrological response were used.

3.2 Results and interpretation

This section illustrates the results of the correlation analysis aimed to identify influencing factors that control the spatial variability of streamflow signatures; Section 3.2.1 ~~shows~~presents the ~~spatial~~ results of the selection of meaningful statistics; Section 3.2.2 identifies climate and landscape indices controlling streamflow signatures and presents consequences for model development.

3.2.1 Selection of meaningful streamflow signatures, climatic indices, and catchment indices

The streamflow signatures defined in Sect. 3.1.1 were calculated for each subcatchment and the values are shown in Table 2 together with the coefficient of variation. All the signatures have a coefficient of variability ~~of the indices, the correlation and regression analysis and~~ bigger than the threshold value of 5%, with the most variable signature being ζ_{LQF} (71%) and the least variable ζ_{HQD} (6%). Therefore, none of these signatures was discarded.

Figure 2 shows the correlations between the streamflow signatures: the lower triangle contains the Spearman's rank correlation and the upper triangle the p-value associated with the correlations. Based on correlations and on its interpretation ~~is presented in Sect. 3.2.2.~~, a subset of ζ can be defined as follows:

~~3.2.1 Spatial and temporal variability of catchment indices~~

~~In Fig. 2, each boxplot shows the variability (between years) of the observed streamflow signatures. This analysis suggests that, based on the signatures ζ_Q, ζ_{RC}, ζ_{BFI}, and ζ_{FI}, the subcatchments can be qualitatively divided in three separate groups:~~

- ~~subcatchments in the north- west hilly part (Frauenfeld and Wängi) characterized by on average lower values of ζ_Q (less than about 700 mm yr^{-1}), ζ_{RC} (less than 0.60), and ζ_{FI} (about 0.30), and higher values of ζ_{BFI} (about 0.50);~~
- ~~subcatchments in the south mountainous part (Appenzell, Jonschwil, Mogelsberg, Mosnang, and St. Gallen) that present completely opposite behaviour in terms of signatures values compared to the first group, with higher ζ_Q (larger, on average, than 1100 mm yr^{-1}), ζ_{RC} (larger than 0.70), and ζ_{FI}, (with the average larger than 0.40), and lower ζ_{BFI}(around 0.40);~~
- ~~subcatchments with intermediate behaviour (all the others, i.e. Andelfingen, Halden, and Herisau) that express a regime that is in between the other two groups.~~

~~This combination of signatures values suggests that the subcatchments in the north- west hilly part, since they have more baseflow and a lower ζ_{FI}, manifest a hydrograph that is more regular (with less short-term variations) than the other two groups.~~

~~Based on the half streamflow period (ζ_{HSP}), the subcatchments can be classified in different groups: Frauenfeld, Wängi, and Mosnang, that on average, reach the 50 % of the total streamflow around the 45 % of the year (February); Jonschwil, St. Gallen, and Appenzell that reach this threshold around the 60 % of the year (April); all the others that have an intermediate behaviour. All the subcatchments present some outliers. These outliers can be explained based on the temporal variability of~~

~~precipitation, which can determine that in some years the 50 % of the total streamflow is reached much earlier than on average.~~

~~Figure 3 illustrates the same analysis, but for the meteorological indices. It is possible to observe in the precipitation (Fig. 3a) and in the aridity index (Fig. 3c) the same patterns that are present in the streamflow (Fig. 2a); nevertheless, the ratio between streamflow and precipitation is not constant, as shown by ζ_{RC} (Fig. 2b), which is higher for wetter subcatchments. Precipitation, varies significantly from subcatchment to subcatchment, with an average value that has a range of variability of more than 500 mm yr^{-1}. PET instead is generally more stable from subcatchment to subcatchment. The average value oscillates of about 50 mm yr^{-1}, with the exception Appenzell, the catchment with highest altitude, where PET is, on average, 100 mm yr^{-1} lower (Fig. 3b).~~

~~The landscape characteristics of the subcatchments (ξ) are summarized in Fig. 4 (refer to Table 1 for the features that cannot be expressed as areal fraction, e.g. ξ_{TE}, ξ_{TSm}, etc.). All the subcatchments present the same aspect (ξ_{TAs}, ξ_{TAew} and ξ_{TAn}), mainly north (30 %). Topography (ξ_{TSm} and ξ_{TSs}), soil characteristics (ξ_{SD} and ξ_{SM}), and geology (ξ_{GA}, ξ_{GC}, and ξ_{GU}) present a high variability, with a difference, between subcatchments, around 40 %. The land use (ξ_{LC}, ξ_{LP}, ξ_{LF}, and ξ_{LU}) is relatively uniform (variation lower than 10 % between catchments) with the only large (but still limited) difference in the urbanized areas. There are also landscape characteristics (e.g. ξ_{LC} or ξ_{LU}) that present a limited coverage over the catchment, lower than 15 %.~~

~~Reducing the time scale from annual to monthly averages it is possible to note differences in seasonality between subcatchments. Figure 5 illustrates the variability of streamflow, precipitation and PET; each line represents the normalized (divided by the average annual value) average value through the years for a subcatchment. Although the (normalized) meteorological variables present similar seasonality between the subcatchments, the streamflow shows stronger variability between subcatchments. Based on streamflow seasonality, the subcatchments can be visually divided in two separate groups: subcatchments that have highest streamflow between October and March (Wängi, Frauenfeld, and Mosnang) and subcatchments that present highest streamflow during late spring and summer (particularly evident in Appenzell). These dynamics are similar to the ones captured by ζ_{HSP}, that shows that the catchments that reach earlier the 50 % of the streamflow are the same that have the highest streamflow between October and March.~~

~~3.2.2 Influencing factors on streamflow signatures~~

- ~~Table 2 shows the correlation coefficients calculated between~~ ζ_{Q} , ζ_{RR} and ζ_{EL} are strongly correlated ($r > 0.72$). We retained ζ_{Q} and discarded ζ_{RR} and ζ_{EL} because both contain climatic information (precipitation) in their definition;
- ζ_{BFI} and ζ_{FDC} are strongly correlated ($r = -0.77$). We decided to retain ζ_{BFI} as it is of easier interpretation (it is a proxy for the importance of groundwater flow, which is a potentially important process for the subsequent model development);

- ζ_{HFD} was kept because it measures the seasonality of the streamflow. Note that ζ_{HFD} is strongly correlated with ζ_{Q} ($r = 0.88$). However, they reflect different properties of the hydrograph. In particular, ζ_{HFD} can be an useful indicator for the effect of snow-related processes;
- ζ_{Q5} and ζ_{HQD} were retained because they have low correlation ($r < 0.71$) with the other selected signatures and because the first represents low flows and the second high flows;
- ζ_{Q95}, ζ_{HQF}, ζ_{LQD}, and ζ_{LQF} were discarded because they all show correlations with the selected signatures.

In summary, the original set of streamflow signatures was reduced to a set of five meaningful signatures, which will be in the subsequent analyses: average daily streamflow (ζ_{Q}), baseflow index (ζ_{BFI}), half streamflow period (ζ_{HFD}), 5$^{\mathrm{th}}$ percentiles of the streamflow (ζ_{Q5}), and duration of high-flow events (ζ_{HQD}).

In terms of climatic indices, Table 3 shows their values together with the coefficient of variation. It can be seen that there are some indices that show very little or no variation at all and, therefore, they could be already excluded from the subsequent correlation analysis; they are: ψ_{HPD} (1 %), ψ_{HPS} (0 %), ψ_{LPF} (4 %), ψ_{LPD} (3 %), and ψ_{LPS} (0 %).

Fig. 3 shows the correlation between the remaining indices. It can be observed they all have strong internal correlation ($r > 0.71$). For this reason it was decided to retain only ψ_{P} and ψ_{FS}, as they have lower correlation. The first represents an important term of the water budget, the latter captures snow dynamics.

Table 4 shows the values of the catchment characteristics considered in this study. All of them have a coefficient of variation larger than the minimum threshold of 5%. Therefore, none of them was excluded based on this criterion. The second criterion for the pre-exclusion of the catchments characteristics, consisting in removing ξ occupying less than 5% of the subcatchments, led to the suppression of ξ_{LC} (which occupies 4% of the subcatchment).

Figure 4 shows the correlations between catchment characteristics; in many cases the high correlation is due to the fact that many indices are complementary (e.g. different types of geology). The following ξ were selected (one index per class):

- ξ_{A} because it is low correlated to the other features;
- ξ_{TE} and ξ_{TAs} in representation of the topography;
- ξ_{LF} for the land use;
- ξ_{SD} representing the soil characteristics;
- ξ_{GC} for the geology.

In summary, the original set of catchment indices was reduced to a set of 5 indices.

3.2.2 Selection of controlling factors on streamflow signatures

Fig. 5 reports the results of the Spearman correlation between climatic indices plus catchment characteristics on streamflow signatures. The upper panel contains the Spearman's rank coefficients and the lower panel shows p-values associated with them.

The following results can be noted:

- The three statistics average precipitation (ψ_{P}), fraction of snow (ψ_{FS}), and average elevation (ξ_{TE}) correlate strongly with average streamflow (ζ_{Q}) and seasonality (ζ_{HFD}) ($r > 0.64$ and p-value< 0.05). This correlation can be interpreted as follows: subcatchments with high elevation (ξ_{TE}) tend to have higher precipitation (ψ_{P}) due to orographic effects, which leads to higher streamflow (ζ_{Q}). They also tend to have more snow (ψ_{FS}) due to lower temperatures, which influences the seasonality (ζ_{HFD}).
- There are then some catchment characteristics that have no correlation ($r < 0.45$) with the streamflow signatures (catchment area (ξ_{A}) and land use (ξ_{LF})) or limited correlation (aspect (ξ_{TAs}) and deep soil (ξ_{SD}), with $r < 0.64$).
- The consolidated geology (ξ_{GC}) presents a strong correlation ($r = -0.87$) only with the baseflow index (ζ_{BFI}) that it is not captured by the other indices.

The streamflow signatures ~~and meteorological and landscape characteristics of the subcatchments; both Pearson and Spearman's rank correlations are reported and the values that are statistically significant (p-value < 0.05) are marked in bold. The two coefficients provide comparable results and, in particular, only a few correlations that are statistically significant according to the Spearman's rank correlation are~~ of low and high flows (ζ_{Q5} and ζ_{HQD}) cannot be explained by any index, with little correlation only with ψ_{P} and ξ_{TE} ($r < 0.60$) that is not ~~considered significant by the Pearson correlations and, in these cases, their p-value is close to 0.05. For this reason, the Pearson coefficient is considered to be a good metric to detect correlation in this case and there is, therefore, no need to look for non-linearity in the following of this analysis.~~

~~Looking at the statistically significant correlations, without any consideration regarding the presence of a causality link, some characteristics can be already excluded from the remaining of the analysis since they are never correlated with the signatures; they are the subcatchment area (ξ_A), the aspect (ξ_{TAs}, ξ_{TAew}, and ξ_{TAn}), and forest land use (ξ_{LF}). Other correlations can be excluded because, while having in some cases a low p-value, the landscape characteristic covers a limited portion of the subcatchment; this is the case of crops and urban land use (ξ_{LC} and ξ_{LU}).~~

~~The remaining correlations are then analysed and accepted (or rejected) considering their relation with other characteristics and the presence (or absence) of a causality link. Analysing ζ_Q, for example, it can be seen from Table 2 that it is, in a decreasing order, correlated to: ψ_P (0.97), ψ_{AI} (-0.97), ξ_{TSm} (0.97), ξ_{TE} (0.97), ξ_{TSs} (0.94), ξ_{SM} (-0.89), ψ_{PET} (-0.87), ξ_{GA} (-0.80), ξ_{GC} (0.80), ξ_{GU} (-0.76), and ξ_{SD} (-0.73). Out of these 11 correlations, some of them can be excluded because either they represent the same feature or because they are linked with other characteristics that are more explanatory. In particular, ξ_{TSm} and ξ_{TSs} are highly correlated and they both represent the topography so only the slope (ξ_{TSm}) is kept. ξ_{SD} and ξ_{SM} represent both the same feature (soil depth) and the same happens for the three types of geology (that are complementary): for these reasons only ξ_{SM} and ξ_{GC} are kept for following analyses. ξ_{TE} and ψ_P are highly correlated and, as explained in Sect. 3.1.2, only the true driver (ψ_P) is considered. The aridity index, finally, is a function of ψ_P and ψ_{PET}. Since ψ_P is already considered, only ψ_{PET} is included in the following analyses.~~

~~The same arguments can be used also to prune the list of the statistically significant correlations for the other signatures, keeping only the most correlated ξ of each landscape characteristic (the first letter of the subscript of each index denotes the~~

~~landscape characteristic that it represents). The outcomes of the causality analysis are reported in Table 2, where the correlations that show a causality link and are not redundant are underlined.~~

~~The selected ψ and ξ are then used, in a decreasing order of correlation (the absolute value is used), in the regression analysis. This is presented in Table 3, where each sub-table contains the results of the regression between the meteorological and landscape characteristics and a signature; the evaluation metrics, r^2 and RSS, are also reported. The following conclusions can be drawn:~~

- ~~in all the cases the null hypothesis (none of the features is necessary to explain the signatures) is rejected;~~
- ~~excluding ζ_{BFI}, the first feature used in the regression is~~ sufficient to reach ~~an acceptable value of the metrics and, thus, the use of the other characteristics does not increase significantly the performance of the model;~~a p-value lower than 0.05.
- ~~when geological and topographical characteristics are used in conjunction, the evaluation metrics do not increase appreciably. This is probably due to the fact that they are redundant, since they present the same spatial distribution across the subcatchments;~~
- ~~regarding the ζ_{BFI}, it can be noticed that, adding information related to the land use ξ_{LP}, the metrics improve substantially (+8 % for r^2 and −60 % for RSS).~~

These ~~conclusions form the basis to formulate modelling guidance (Sect. 3.3).~~

~~**3.3 Guidelines for modelling**~~

~~The~~ results ~~of the correlation and regression analysis are, in this paper,~~ are the premise ~~of the hydrological~~for designing meaningful model experiments.

3.3 Hypotheses for model building~~.~~

Our hypothesis is that only a model that ~~is able to account~~accounts for the influencing factors that affect the streamflow signatures will be able to reproduce spatial streamflow variability. ~~So it is necessary, at~~In this ~~point, to~~section, we synthetize the outcomes of previous analyses in the form of ~~guidelines~~testable hypotheses for model building ~~the model~~.

1. The precipitation ~~and, in general, all the meteorological variables are~~ is the first ~~drivers~~driver of the ~~hydrological~~differences in the water balance of the subcatchments. The effect of topographic variability~~; they show a statistically significant correlation with all the signatures and they are the first control~~ manifests itself primarily as an influence on ζ_Q~~. For this reason, the hydrological model should be able to distribute the inputs across the subcatchments; doing this, the model is automatically incorporating information about the~~ precipitation (amount and type). Accounting for variability of precipitation therefore implicitly reflects such effect of topography ~~(specifically~~on the ~~elevation)~~hydrograph, since ~~all~~ the inputs were interpolated taking into account the effect of the elevation (Sect. 2).

~~2. The signature ζ_{HSP} and the different monthly patterns between streamflow and meteorological inputs show a seasonal effect due to the presence of snow. Therefore the model should be able to use the temperature (that is distributed across the subcatchments) to separate the precipitation between snow and rainfall and to reproduce the melting process.~~

2. ~~Out of all the landscape~~Snow related processes (e.g. amount of snow, timing of snowmelt) control differences in streamflow seasonality between subcatchments.

3. Geology exerts an important control on the partitioning between quick flow and baseflow.

4. The other catchment characteristics~~, the ones that are more correlated are topography and geology; since they also have a similar spatial distribution, only geology should be kept because it has, overall, a higher correlation~~ (e.g. soil, vegetation) show little or no correlations with the streamflow signatures~~. There are also other characteristics that have proven to be related with some signatures (e.g. land use with ζ_{BFI} and ζ_{FI} or soil depth with ζ_{BFI}) but, for the sake of keeping~~ and therefore they should not be considered if the idea is to keep the model as simple as possible~~, they~~.

~~3.~~ These hypotheses will ~~not~~ be ~~considered.~~ ~~This guidance will result into~~tested through specific model ~~decisions~~comparisons, described in Sect. ~~4.1.1 and in the selection of the model experiments of Sect.~~ 4.1.5.

4 Modelling

4.1 Methods

This section describes the approach for building and testing a ~~semidistributed~~semi-distributed hydrological model designed to represent the observed streamflow and particularly the observed spatial variability of streamflow signatures. The ~~modelling choices are~~general model structure is explained in Sect. 4.1.1 ~~and follow the discussion of the regression analysis presented in Sect. 3.3;~~, the error model and the calibration procedure are described in Sect. 4.1.2 and 4.1.3, the metrics utilized to assess the performance are shown in Sect. 4.1.4, and the model experiments done are illustrated in Sect. 4.1.5.

4.1.1 General structure of the hydrological model

~~The need to provide simultaneous streamflow predictions at the various points within the Thur catchment requires at a minimum a subdivision into subcatchments based on the location of the gauging stations. In the spirit of semidistributed modelling, a common way to account for the spatial heterogeneity of hydrological behaviour is to consider a variable number of HRUs (see Sect. 4.1.5). Portions of the entire catchment belonging to the same HRU are supposed to have the same hydrological behaviour and, for this reason, are described by the same model structure and parameters set (but can have different states because of spatial variability of the forcings). Specific choices for the HRUs, motivated by the results of the regression analysis (Sect. 3.3), are described in Sect. 4.1.5.~~

~~The results of the regression analysis have indicated that~~ We describe here the general model structure. Specific choices for the various experiments are described in Section 4.1.5. The model uses a two-layers decomposition of the catchment:

1. Subcatchments are defined by the presence of the gauging stations; this subdivision was due to the necessity of having locations in the model where the streamflow was both observed and simulated and, therefore, it was possible to calibrate and evaluate the parameters of the hydrological model. This layer of decomposition was used for the distribution of the meteorological inputs (precipitation~~ is a dominant control on average streamflow (Sect. 3.3, point 1). Therefore, precipitation needs to be distributed at least at the level of subcatchments. As a result, portions~~, PET, temperature), that are aggregated at the subcatchment scale.
2. HRUs are defined based on catchment characteristics (e.g. topography, geology or vegetation); they represented parts of the catchment ~~belonging to the same HRU but located in different~~that are supposed to have a similar hydrological response to the meteorological forcing. Each HRU is characterized by its own parameterization. Different definitions of HRUs were tested, as described in Section 4.1.5.

Each HRU has a unique parameterization. However, given the choice of discretizing the inputs per subcatchment, a HRU that spans multiple subcatchments will generally have different states~~. Using the same terminology of Fenicia et al. (2016), the smallest landscape units in which the catchment is discretized are called 'fields'. In this case, the total number of 'fields' is obtained by summing the total number of HRUs present~~ in each subcatchment. Therefore, the same HRU needs its own model representation in each subcatchment where it is present. For more details about our model implementation of "HRUs" refer to Fig. 4 of Fenicia et al. (2016).

The model was built using the modelling framework SUPERFLEX (Fenicia et al., 2011). ~~We have chosen~~In contrast to Fenicia et al. (2016), for simplicity we chose a unique structure to represent the various HRUs (as said above, this structure will generally have different parameters in order to represent the hydrological behaviour of distinct HRUs). The structure used to represent the HRUs is represented in Fig. 6 with the equations listed in the Appendix A. ~~Because of the importance of snowmelt in controlling streamflow seasonality (Sect. 3.3, point 2), the~~The structure includes a snow reservoir (WR~~).~~), with inputs distributed per subcatchments. Snowmelt and rainfall are input to an unsaturated reservoir (UR), which determines the portion of precipitation that produces runoff. This flux is split through a fast reservoir (FR), designed to represent the peaks of the hydrograph, ~~preceded~~proceeded by a lag function to offset the hydrograph, and a slow reservoir (SR), designed to represent baseflow. This structure ~~is~~was chosen to be parsimonious while general enough to reproduce typical hydrograph behaviour; it was tested in previous applications (e.g., van Esse et al., 2013; Fenicia et al., 2014; Fenicia et al., 2016) demonstrating its suitability to reproduce a wide range of catchment responses. It also resembles popular conceptual hydrological models such as HBV (Lindstrom et al., 1997) and HyMod (Boyle, 2003), which are shown to have wide applicability.

4.1.2 Error model

As commonly done in hydrological modelling (e.g., McInerney et al., 2017), we here account for uncertainties by considering a probabilistic model of the observations $\boldsymbol{Q}(\boldsymbol{\theta},\boldsymbol{x})$, where $\boldsymbol{\theta}$ is the vector of parameters and $\boldsymbol{x}$ the model input, which is composed of a deterministic hydrological model $\boldsymbol{h}\cancel{(\boldsymbol{\theta_h},\boldsymbol{x})}(\boldsymbol{\theta}_{\mathrm{h}},\boldsymbol{x})$ (illustrated in Sect. 4.1.1) and a random residual error term $\boldsymbol{E}(\cancel{\boldsymbol{\theta_E}}\boldsymbol{\theta}_{\mathrm{E}})$ that accounts for all data and model uncertainties ($\cancel{\boldsymbol{\theta_h}}\boldsymbol{\theta}_{\mathrm{h}}$ and $\boldsymbol{\theta}_{\cancel{E}\mathrm{E}}$ represent the hydrological and the error parameters):

$$z[\boldsymbol{Q}(\boldsymbol{\theta},\boldsymbol{x});\lambda] = z[\boldsymbol{h}\cancel{(\boldsymbol{\theta_h},\boldsymbol{x})}(\boldsymbol{\theta}_{\mathrm{h}},\boldsymbol{x});\lambda] + \boldsymbol{E}(\cancel{\boldsymbol{\theta_E})}\cancel{(3}\boldsymbol{\theta}_{\mathrm{E}}) \qquad (\cancel{3}4)$$

where $z[\boldsymbol{y};\lambda]$ ~~represent~~represents the Box–Cox transformation (Box and Cox, 1964) with parameter λ, which is used to account for heteroscedasticity (stabilize the variance). For $\lambda \neq 0$:

$$z\cancel{[\boldsymbol{y};\lambda]}[y_{\mathrm{t}};\lambda] = \cancel{\frac{y^{\lambda}-1}{\lambda}}\frac{y_{\mathrm{t}}^{\lambda}-1}{\lambda} \qquad (\cancel{4}5)$$

The residual error term is assumed to follow a Gaussian distribution with zero mean and variance σ^2

$$\cancel{\boldsymbol{E}\sim}E_{\mathrm{t}} \sim N(0;\sigma^2) \qquad (\cancel{5}6)$$

The error model has, therefore, two parameters (λ and σ^2); the first was fixed to 0.5 (McInerney et al., 2017) and the second was inferred.

This choice of error model (Gaussian noise applied to the Box–Cox transformation of the streamflow) allows for an explicit definition of the likelihood function (McInerney et al., 2017)

$$\cancel{p(\boldsymbol{q_{obs}}|\boldsymbol{\theta_h},\boldsymbol{\theta_E},\boldsymbol{x}) = \prod z'(\boldsymbol{q_{obs}}|\boldsymbol{\theta_E}) f_N(\boldsymbol{E}|0;\sigma^2) \qquad (6)}$$

~~where~~$p(\boldsymbol{q}_{\mathrm{obs}}|\boldsymbol{\theta}_{\mathrm{h}},\boldsymbol{\theta}_{\mathrm{E}},\boldsymbol{x}) = \prod_{\mathrm{t}=1}^{\mathrm{T}} z'(q_{\mathrm{obs,t}}|\boldsymbol{\theta}_{\mathrm{E}}) f_N(E_t|0;\sigma^2) \qquad (7)$

where T represents the length of the time series, f_N is the Gaussian probability density function (PDF) and $z'(\boldsymbol{q_{obs}}|\boldsymbol{\theta_E})$ is the derivative of $z(\boldsymbol{q_{obs}},\boldsymbol{\theta_E})$ with respect to $\boldsymbol{q}$ evaluated at the observed data $\boldsymbol{q_{obs}}$. Specifying Eq. (~~6~~7) for the case where $z(\boldsymbol{q_{obs}};\boldsymbol{\theta_E})$ is defined by Eq. (~~4~~5), the expression of the likelihood function becomes:

$$\cancel{p(\boldsymbol{q_{obs}}|\boldsymbol{\theta_h},\boldsymbol{\theta_E},\boldsymbol{x}) = \prod \boldsymbol{q_{obs}}^{(\lambda-1)} f_N(\boldsymbol{E}|0;\sigma^2) \qquad (7)}$$

$$p(\boldsymbol{q}_{\mathrm{obs}}|\boldsymbol{\theta}_{\mathrm{h}},\boldsymbol{\theta}_{\mathrm{E}},\boldsymbol{x}) = \prod_{\mathrm{t}=1}^{\mathrm{T}} q_{\mathrm{obs},t}^{(\lambda-1)} f_N(E_t|0;\sigma^2) \qquad (8)$$

Equation (8) represents the likelihood function that is then used, together with an uniform prior distribution, to calibrate the parameters of the model as described in Sect. 4.1.3.

4.1.3 Calibration

Parameter calibration was performed ~~by optimizing~~with the ~~parameters~~objective of ~~the~~maximizing their posterior ~~distribution~~density. According to Bayes equation, the posterior distribution of model parameters is expressed as the product between the prior distribution and the likelihood function; since an uniform prior was used for the parameters, this is equivalent to maximizing the likelihood function in the defined parameter space; the optimization procedure was ~~done using~~performed with a multi–start quasi–Newton method (Kavetski et al., 2007) with 20 independent searchers. We empirically established that with models of our complexity (about 10 parameters), 20 independent searches provide good confidence that a global optimum is found.

The evaluation of the model ability to reproduce streamflow was carried out in space–time validation~~.~~ (see also Fenicia et al., 2016). For this purpose, the time domain was divided in two periods of 12 years each (from 01 September 1981 to 01 September 1993, and from 01 September 1993 to 01 September 2005) and the subcatchments were split ~~in~~into two groups (A and B), according to a spatial alternation (subcatchment in group A flows into a subcatchment in group B that flows into one in group A and so on); the subcatchments belonging to group A are Andelfingen, Herisau, Jonschwil, St. Gallen, Wängi and the ones in group B are Appenzell, Frauenfeld, Halden, Mogelsberg, Mosnang. This method implies a division of the space–time domain in four ~~parts~~quadrants, such that the model can be calibrated in one quadrant and validated in the other three. For space–time validation, the model was calibrated using each group of subcatchment and each period, and validated using the other group of subcatchment and period. That is, the model calibrated using group A and period 1 was validated using group B and period 2, and so on for the other 3 combinations of subcatchments and groups. The model output in the 4 space–time validation periods was then combined, to calculate model performance using various indicators (see Sect. 4.1.4). Results are presented for space time validation, which represents the most challenging test of model performance.

4.1.4 Performance assessment

Model performance was assessed using the following metrics:

1. Time series metrics, which evaluate the ability of reproducing streamflow time series. The metrics used for this assessment are the following:
   - Normalized log–likelihood (LL), that is~~,~~ the logarithm of Eq. (~~7~~8) normalized by the number of time steps present in the time series. This metrics corresponds to the objective function used for model optimization. It can be observed that, since λ is fixed at 0.5 in the Box–Cox transformation, model calibration is equivalent to maximising the Nash–Sutcliffe efficiency (NS) calculated with the square root of the streamflow. LL is not bounded but a higher value means a better match between two time series since, in this case, the absolute value of the residual is smaller and, thus, their PDF higher.
   - Nash–Sutcliffe efficiency

$$NS\cancel{(\boldsymbol{q_{obs}},\boldsymbol{q_{sim}}) = 1 - \frac{\sum_{t=1}^{T}\left(q_{sim}^{t}-q_{obs}^{t}\right)^{2}}{\sum_{t=1}^{T}\left(q_{obs}^{t}-\overline{q_{obs}}\right)^{2}}}$$

$$\cancel{(8}(\boldsymbol{q}_{\mathbf{obs}},\boldsymbol{q}_{\mathbf{sim}}) = 1 - \frac{\sum_{t=1}^{T}\left(q_{\mathrm{sim,t}}-q_{\mathrm{obs,t}}\right)^{2}}{\sum_{t=1}^{T}\left(q_{\mathrm{obs,t}}-\overline{\boldsymbol{q}_{\mathbf{obs}}}\right)^{2}} \qquad (9)$$

Which is often used in hydrological applications, and it provides a sense of general quality of the simulations. NS is bounded between $-\infty$ and 1, with 1 meaning a perfect match.

2. Signature metrics, which determine the ability of reproducing the selected streamflow signatures (ζ) ~~presented~~which, as illustrated in ~~the regression analysis part (Sect.~~Section 3.2.1~~.1), that is, of~~ , are average daily streamflow (~~ζ_{Q}), runoff coefficient (ζ_{RC}~~ζ_{Q}), baseflow index (~~ζ_{BFI}), flashiness index (ζ_{FI}), and~~ ζ_{BFI}) mean half streamflow ~~period~~date (ζ_{HFD}), 5^{th} percentile of the streamflow (ζ_{Q5}), and duration of high-flow events (~~ζ_{HSP}~~ζ_{HQD}). The accordance between simulated and observed signatures was assessed both visually and using the ~~Pearson~~Spearman's rank correlation.

~~This set of metrics, together with the fact that they are calculated in space time validation (Sect. 4.1.3), provides a comprehensive assessment of model performance.~~

The use of multiple metrics for assessing model performance enables a comprehensive assessment of various characteristics of the simulations. Time series metrics were designed to appraise the general quality of the model fit. Signatures, instead, were designed to highlight selected characteristics of the data at the expense of others.

4.1.5 Model experiments for testing the results of the correlation analysis

Using the model structure described in Sect. 4.1.1~~.1, several~~, four model ~~variants are~~configurations were compared~~. The main motivations for such comparisons are as follows:~~

- ~~verify that models that account for the influencing factors identified through the regression analysis indeed lead to an improved representation of streamflow spatial variability;~~
- ~~provide a mechanistic interpretation of how influencing factors affect streamflow, which cannot be achieved~~ by ~~regression analysis;~~
- ~~get some insights on the relationship between model complexity and performance.~~

~~The model variants and their specific rationale are described below:~~

- ~~In order to verify the effect of spatial distribution of landscape properties, we constructed a reference model with a single HRU, called M1, (i.e. no spatial distribution of landscape properties); in this case only the input variability (the catchment is still divided in subcatchments) is considered.~~
- ~~In order to verify that geology controls streamflow variability (see Sect. 3.3, point 3), particularly by influencing baseflow conditions, a two HRUs model, called M2, was implemented, dividing~~varying the ~~three geology classes in unconsolidated, for~~ number and the ~~first HRU, and consolidated and alluvial, for the second HRU (see Sect. 2 and Fig. 1d).~~

~~In order to verify that eventual improvements in performance brought by M2 compared to M1 are not just due to increase in complexity, we implemented a two~~ definition of the HRUs, and changing the structure of the HRUs ~~model, called M3, using the land use to discretize the domain. The land use classes were arbitrarily defined so that the first HRU contains forest and~~

~~crops and the second occupies the rest of the catchment.~~ (Fig 6). The objective of the experiments was to test the hypotheses 1-4 in Sect. 3.3.

The first hypothesis (precipitation controls the water balance) is tested with the model M0, with uniform parameters on the catchment (i.e. a single HRU) and distributed precipitation input. This model does not consider snow processes. We expect that this model will be able reproduce differences in streamflow averages between subcatchments.

The second hypothesis (snow controls seasonality) is tested with the model M1. Relatively to M0, M1 accounts for snow processes, represented by simple degree day snow module (see Kavetski and Kuczera, 2007), with inputs (temperature) distributed per subcatchment.

The third hypothesis (geology controls baseflow) is tested with the model M2. Relatively to M1, M2 considers two HRUs, defined based on geology type. One HRU contains the areas with consolidated geology while the other HRU contains the rest of the catchment (unconsolidated and alluvial geology together).

The fourth hypothesis (other catchment characteristics should not be considered if the idea is to keep the model as simple as possible), is exemplified by the model M3. M3 is analogous to M2 except that HRUs are based on catchment characteristics that did not show correlation with the streamflow signatures. Among those characteristics, we have selected land use, and considered an HRU based forest and crops and the second one that occupies the rest of the catchment.

The total number of the calibrated parameters depends on the number of HRUs and on the structure used to represent them: it was ~~nine in the first experiment (Table A1)~~8 for M0, 9 in M1, and 13 in ~~the other two~~M2 and M3, where ~~five~~5 parameters were linked between different HRUs~~;~~ (Table A1); those parameters are: $\cancel{C_e}C_{\mathrm{e}}$ that governs the evapotranspiration, $\cancel{t_{rise}^{OL}}t_{\mathrm{rise}}^{\mathrm{OL}}$ and $\cancel{t_{rise}^{IL}}t_{\mathrm{rise}}^{\mathrm{IL}}$ that control the routing in the river network, $\cancel{k_{WR}}k_{\mathrm{WR}}$ that regulates the outflow of the snow reservoir, and $\cancel{S_{max}^{UR}}S_{\mathrm{max}}^{\mathrm{UR}}$ that determines the behaviour of the unsaturated reservoir.

4.2 Results and interpretation

This section presents the results of the modelling experiments. Section 4.2.1 illustrates model results in terms of hydrograph metrics. Section 4.2.2 presents model results in terms of signatures. An interpretation of the results, including a comparison with the conclusions of the ~~regression~~correlation analysis, is given in Sect. 4.2.3.

4.2.1 Model performance in terms of hydrograph metrics

Figure 7a shows the values of the likelihood function (corresponding to the calibration objective function) for the ~~three~~four models in calibration and validation. It can be observed that M0 is, by far, the worst model, having a low value of likelihood. Moving to the other three models, it can be seen that, during calibration, M1, which has the lowest number of calibration parameters, has the lowest ~~likelihood value of the three, indicating lowest~~ performance, whereas M2 and M3 have similar higher likelihood values. This behaviour ~~continues~~persists in time validation, with M2 and M3 that outperform M1. In space and space–time validation, however, M3 has the lowest likelihood value of the three, whereas M1 and M2 limit their decrease in performance, ~~having~~ranking, respectively, ~~the~~ second and the first ~~value~~in terms of optimal likelihood value.

The likelihood function represents an aggregate metric of model performance; in order to get a sense of appreciation of model fit on individual subcatchments, Fig. 7b reports the values of Nash Sutcliffe efficiency in space time validation for each of the subcatchments. On average, M2 has the best performance of all models (NS = 0.79), followed by M1 (NS = 0.78~~)~~ ~~and~~), M3 (NS = 0.77~~).~~), and M0 (NS = 0.68). M3 ~~has~~and M0 have the highest variability of performance, with NS values between 0.58 and 0.86 and between 0.59 and 0.81. M1 and M2 have similar spread of NS values, ranging from 0.69 to 0.85 for M1 and from 0.73 to 0.87 for M2. Therefore, M1 and M2 have a more stable performance across subcatchments than M3. M3 obtains a significantly worse performance than the other 2 models on Mosnang, where it reaches a NS value of 0.58 (M1 and M2 have values of 0.69 and 0.73 respectively).

It can also be observed that M2 is generally better than M1, with NS values that are higher or approximately equal except for the subcatchments Andelfingen and Halden, where the NS is slightly worse (however still higher than 0.80). M3 is clearly better than M1 on Andelfingen, Frauenfeld and Wängi, and clearly worse on Herisau and Mosnang. In particular, in Mosnang (the smallest basin), M3 reaches the worst performance of all models on all subcatchments.

Regarding M0, it is interesting to observe that it has the worst performance (among all the subcatchments) in Appenzell, which is the subcatchment that is most affected by snow ($\psi_{\mathrm{FS}} = 0.21$), while it reaches a performance similar to M1 in Frauenfeld and Wängi, which are two subcatchments with almost no snow.

4.2.2 Model performance in terms of signature metrics

Figure 8 compares the ~~observed~~ability of M0 and ~~simulated signatures for~~ M1~~. Figure 9~~ to capture the signatures representing average streamflow (ζ_{Q}) and ~~Fig. 10 show the same analysis for M2 and M3 respectively.~~seasonality (ζ_{HFD}). The analysis is presented for space–time validation ~~only.~~and, for ζ_{HFD}, it focuses only on the four subcatchments that are most affected by the snow ($\psi_{\mathrm{FS}} > 0.10$) to emphasize the differences between the results of the two models. Each colour represents a different subcatchment and each dot a year; the red dashed line has a 45 ° slope and ~~it is~~represents where the dots should align in case of perfect simulation results. ~~The Pearson correlation coefficient~~The Spearman's rank score is also reported and gives information about the degree of ~~linear~~ dependency between the two variables. It is important to ~~underline~~stress that the models have not been calibrated using any of ~~these~~the signatures as objective function~~ and,~~, which therefore~~, they~~ represent an independent evaluation metric.

~~M1 (Fig. 8)~~It can be observed that M0 represents ~~relatively accurately ζ_Q,~~ζ_{Q} as well as M1, with almost no difference between the two models. Focusing on the ability of capturing ~~ζ_{RC}, and ζ_{HSP}~~ (ζ_{HFD}, it can be seen that with M0 the points all lie in the upper-left part of the plot, meaning that this model underestimates the signature values. With respect to M1, instead, the points are more aligned around the diagonal. This difference in performance is also exemplified by the value of r that is 0.~~96, 0.83,~~66 for M0 and 0.85 for M1.

Figure 9 compares the observed and ~~0.92 for ζ_Q, ζ_{RC},~~simulated signatures for the other three models (M1, M2, and ~~ζ_{HSP}~~M3). All of them are extremely good in representing ζ_{Q} (r is 0.95, 0.96, and 0.95 for M1, M2, and M3 respectively~~). On the other hand, the model shows clear~~) and ζ_{HFD} (r is 0.88, 0.88, and 0.87 for M1, M2, and M3 respectively). In all cases the cloud of points appears aligned to the diagonal meaning that the three models are able to capture the value of the signatures each year. Moreover, there is no sensible difference in the various models in representing those signatures.

The performance of all the models decreases for ζ_{Q5} where the models have a similar performance with r equal to 0.62, 0.66, and 0.61 for M1, M2, and M3 respectively. The points cloud is still aligned to the diagonal but it is quite dispersed, especially if compared with ζ_{Q} and ζ_{HFD}, meaning that the models capture the general tendency but have deficiencies ~~in matching~~capturing the inter-annual variability.

In terms of ζ_{BFI}, M2 performs clearly better than the other ~~signatures (r is 0.20 and 0.37 for ζ_{BFI} and ζ_{FI} respectively): the points cloud is quite dispersed meaning that the~~ models. It is the only model ~~is not able to capture the variability between years and subcatchments. Moreover,~~ that is able to represent this signature, with r equal to 0.83 and the points that align to the diagonal. The other two models have a lower performance (r equal to 0.31 and 0.52 for M1 and M3 respectively) with the points cloud that is quite dispersed and the dots align~~ (regression line not plotted to avoid an overcrowded plot)~~ almost vertically, implying that the simulated values have a range of variability that is definitely smaller than the ~~measured~~observed data.

~~M2 (Fig. 9) has a performance similar to M1 in terms of the signatures ζ_Q, ζ_{RC}, and ζ_{HSP} (r is 0.96, 0.88, and 0.92 for ζ_Q, ζ_{RC}, and ζ_{HSP} respectively). However, in terms of ζ_{BFI}, the representation is much better than M1 (r equal to 0.83 vs 0.20 for M1), with the points cloud that is more aligned on the diagonal, indicating not only correlation but also a good match of the absolute values. Compared to M1, finally, ζ_{FI} presents a much better alignment (r is 0.88), but the points are still far from the diagonal, indicating a poor agreement of the absolute values; in particular, the model tends to consistently underestimate ζ_{FI}.~~

~~M3 (Fig. 10), finally, has a performance that is in between the other two models; the representation of the signatures ζ_Q, ζ_{RC}, and ζ_{HSP} is similar to the one achieved with the other two models, with r that is equal to 0.96, 0.84, and 0.90 respectively. On the other hand, the model misrepresents the other signatures (r is 0.46 and 0.44 for ζ_{BFI} and ζ_{FI} respectively): the point clouds, in this case, are more similar to the one obtained with M1, that is, they are more dispersed and aligned almost vertically.~~

Figure 10 shows the comparison between observed and simulated ζ_{HQD}; since this signature requires a long time window to be computed, it is not calculated year by year (as done with the other signatures) but it is available only the aggregated value over the 24 years. The performance of M1 and M2 is overall good, with r that is 0.77 and 0.69, while M3 shows some deficiencies (r equal to 0.48); all the models tend to slightly overestimate the duration of high flow events with most of the points that lie on the right side of the diagonal.

4.2.3 Interpretation of hydrological model results

The results of the hydrological model experiments ~~indicate~~appear to support our hypothesis that ~~accounting~~only models that account for the influencing factors ~~identified through~~that affect the ~~regression analysis indeed lead to an improved representation of~~ streamflow signatures are able to reproduce streamflow spatial variability~~. The~~ (see Sect. 3.3). This provides confidence that those models are a realistic representation of dominant processes in the catchment.

In particular, the results of ~~M1~~M0 show that accounting for the spatial heterogeneity of the ~~inputs~~precipitation alone is sufficient to achieve ~~to~~ a good accuracy signatures of water balance, with r of 0.~~96~~95 for average streamflow ζ_Q. More complex models with more HRUs and more parameters do not result in any improvement in reproducing the average streamflow signature.~~ The same considerations can be made also for ζ_{RC} and ζ_{HSP} that are well represented by all the three models.~~

The differences between M1 and M0 show that differences in streamflow seasonality ζ_{HFD} can be largely attributed to the (spatially variable) effect of snow accumulation and melting. More complex models (M2 and M3) do not demonstrate an improvement in this signature.

M2 determines a large improvement in matching signatures of baseflow variability. The ability of fitting ~~ζ_{BFI}~~ζ_{BFI} goes from 0.~~20~~31 for M1 to 0.83 for M2. This result confirms that geology influences spatial variability of quickflow vs baseflow partitioning, as indicated by ~~regression~~correlation analysis.

M3 reassures that the relatively good results of M2 are not just due to increasing complexity. Although this model performs slightly better than the M1 in terms of matching signatures such as ~~of ζ_{BFI} and ζ_{FI}~~ζ_{BFI}, M2 is still much better (e.g. the ~~Pearson coefficient~~Spearman's rank score for ~~ζ_{BFI}~~ζ_{BFI} is 0.83 for M2 and 0.~~46~~52 for M3).

All the models do not preform particularly well in reproducing ζ_{Q5} and ζ_{HQF}. These problems shows that such models may not represent well extreme values (high and low flow), and therefore they are still amenable to improvements.

Overall, distributing the inputs is sufficient to get good performance metrics, water balance, and seasonality, confirming the fact that the precipitation rate and the partitioning between rainfall and snow are the first controllers on these hydrograph characteristics, but, if we want to capture also other important characteristics of the hydrograph~~ shape~~, described by ~~the other~~ signatures~~,~~, like ζ_{BFI}~~,~~ the discretization of the catchment in HRUs is necessary. This discretization has to be carefully made and a preliminary analysis to understand dominant influencing factors on signatures can help in this decision. As shown in Fig. ~~10~~9, if we use characteristics that are not strongly correlated with the signatures (e.g. land use) the results are worse than if we choose characteristics that show a correlation with signatures (e.g. geology). M2 is capable of capturing the signatures not just because it is more complex than M1, but because it incorporates the causality link between the geology and the streamflow signatures in its structure.

~~The underestimation of ζ_{FI}, which applies to all the models (including M2), has partly to do with the fact that in our analysis observed signatures are compared with the output of the deterministic model without adding the error term. This results in a hydrograph that is more regular that the probabilistic one resulting in a lower ζ_{FI}. Although conceptually it is more~~

~~appropriate to compare the observations to the probabilistic simulations, given that our error model is relatively simple it would have resulted in unrealistic signatures. The development of error models that preserve the signatures of the observations is a subject of current research (e.g., Ammann et al., 2018).~~

5 General discussion

Explaining the spatial variability observed in catchment hydrological behaviour by identifying the most important controls on water fluxes and pathways is a major focus of catchment hydrology and a central theme in classification studies (e.g., McDonnell and Woods, 2004; Wagener et al., 2007). A common approach for interpreting the spatial variability of catchment responses is through ~~regression~~correlation based analyses, which seek correlations between ~~meteorological~~climatic or catchment characteristics and streamflow signatures (e.g., Lacey and Grayson, 1998; Bloomfield et al., 2009). One of the issues with this approach is that correlation does not always imply causality, and the presence of multiple correlated variables can obscure process interpretation.

In this study, we combine ~~regression~~correlation analysis for identifying dominant influencing factors on streamflow signatures with hydrological modelling, by using the interpretation of the ~~regression~~correlation analysis as an inspiration for generating testable model ~~structure design.~~hypotheses. The combination of ~~regression~~correlation analysis on streamflow signatures and hydrological modelling is beneficial because on the one hand, the speculations on dominant processes resulting from the ~~regression~~correlation analyses can be verified in the modelling process. Specifically, we developed model experiments to test the influence of precipitation spatial distribution on streamflow average and seasonality, and the influence of geology on quickflow vs baseflow partitioning. On the other hand, model building benefits from guidance resulting from preliminary ~~regression~~correlation analysis. The construction of a distributed model requires several decisions (e.g., Fenicia et al., 2016), including how to "break–up" the catchment in a meaningful way, and preliminary ~~regression~~correlation analysis can motivate some of these decisions. For example, the HRUs defined based on geology, as suggested by ~~regression~~correlation analysis resulted in better model performance than HRUs based on land use, particularly in the representation of streamflow signatures.

Although several modelling decisions were guided by data analysis, it should be noted that alternative decisions would have been similarly consistent with the data. For example, both precipitation and elevation are correlated with average streamflow, and both geology, topography and soil type characteristics are correlated between each other and with baseflow index (Section 3.2.2 and Figure 5). The correlation of catchment characteristics (e.g. geology, soil and topography) can be attributed to the fact that they evolved together in the shaping of the catchment morphology (e.g. mountainous regions have impervious topography with shallower soil and, for these reason, are less suitable for human activities, influencing land use). The decisions on which variables are chosen to reflect a causality link is not always obvious from correlation analysis alone, and it requires expert judgment, which is not always generally shared.

Our results on the Thur catchment with respect to the effect of meteorological inputs on average streamflow and of the geology on baseflow index are in general agreement with previous work. Kuentz et al. (2017) made a classification study over more than 40000 catchments across all Europe (of which almost 2700 are gauged) and found that the rainfall is the first controller of the average streamflow, geology controls the BFI, topography the ~~FI~~flashiness index, and, for most of the cases, land use is the second controller of them; Bloomfield et al. (2009) used a linear regression model and linked the lithology of the Thames Basin (UK) with the BFI; Lacey and Grayson (1998) noted that geology controls the BFI in two ways, storing the water and impacting the soil formations; Fenicia et al. (2016) compared different model structures and catchment discretization methods in the Attert Basin (Luxemburg) and discovered that the best model was the one that incorporates a spatial representation of the meteorological inputs and of the geology.

On the other hand, this general tendency should not be generalized to all places. For example, Mazvimavi et al. (2005) found that geology was not important for the BFI, as in their case study the aquifer was deep and disconnected from the river. Bouaziz et al. (2018) found a strong influence of regional groundwater flow in the Meuse catchment which altered the water balance.

One of the main limitations of this work is the restricted number of catchments involved and the limited spatial extension of the study. For this reason, it is difficult to generalize the results to other climatic regions. The subcatchments belong all to the same region and the landscape and climatic characteristics, while varying substantially within the basin, can still be quite different from characteristics found elsewhere.

The limited number of catchments involved in this study (only 10) can also pose some problems ~~with~~in the ~~use of techniques such as linear regression with multiple features, which can lead to overfit. Another limitation in the regression study is that~~correlation analysis, where only linear or monotonic correlations have been investigated while other forms of relationship, including the mutual interaction between various influencing factors, have been neglected. This can lead to the exclusion of characteristics that are indirectly related to the streamflow signatures.

6 Conclusion

In this study, we presented a methodology for the construction of a ~~semidistributed~~semi-distributed hydrological model where model ~~decisions~~hypotheses, instead of being made a–priori, are informed by preliminary ~~regression~~correlation analysis on streamflow signatures. Besides providing guidance to model development, the proposed approach is useful in the fact that modelling can be used to test specific hypotheses on dominant processes resulting from ~~regression~~correlation analysis.

Our analysis ~~is~~was applied to the Thur catchment, with the objective of understanding the main controls on streamflow spatial variability. The main findings ~~can be~~are summarized in the following points:

1. ~~there is a~~we found large spatial variability between the subcatchments of the Thur ~~catchment~~ in terms of various ~~characteristics of the hydrographs~~ streamflow signatures reflecting multiple temporal scales: yearly, seasonal and

event scale. In terms of climatic characteristics, indices reflecting fraction of snow, precipitation totals, and aridity varied considerably between catchments. Other precipitation characteristics such as season, frequency and duration of dry and wet days did not vary significantly between catchments. In terms of landscape characteristics, there is large variability of topography (e.g. from upstream mountainous to downstream flat areas), geology (with unconsolidated, more permeable, and consolidated, relatively impermeable formations), and soils (with low depths in the mountains, and large depth in the floodplains) in all catchments;

2. ~~meteorological~~based on correlation analysis and expert judgment, we determined that climatic variables, especially ~~the~~ precipitation average, are the main controls ~~the~~ on streamflow average yearly values ~~and seasonality, whereas they don't have a great effect on the shape of the hydrograph in response to specific rainfall events~~; the fraction of snow is responsible for streamflow seasonality by delaying the release of winter precipitation to the spring season, and geology controls the baseflow index, with a higher fraction of unconsolidated material determining higher baseflow;

- ~~the shape of the hydrograph in response to rainfall events is mainly controlled by the catchment characteristics with the geology that plays the main role. Higher proportion of consolidated material has an influence on the baseflow vs quickflow portioning, causing lower baseflow and higher peaks;~~
- ~~only hydrological models that are able to reflect spatial variability of precipitation and difference in hydrological behaviour between geologies can correctly represent the streamflow signatures considered in this study;~~

3. ~~the causality links found by the regression analysis are used and confirmed by the hydrological model.~~the results of the correlation analysis were translated into a set of model hypotheses: a model with uniform parameters and distributed precipitation input (M0), the addition of a snow component (M1), the subdivision of the catchment in geology based HRUs (M2), and the alternative subdivision the catchment using vegetation based HRUs (M3);
4. using model comparison, and a validation approach that considers model performance (also in terms of signatures) in space time validation, we confirmed that model decisions based on correlation analysis were appropriate. In particular, we confirmed that M0, in spite of a generally poor performance, is sufficient to capture signatures of streamflow average. M1 improves signatures of streamflow seasonality. M2 enables reproducing signatures such as the baseflow index. Model modifications that are not in line with the results of the signature analysis, such as subdividing the catchment using vegetation based HRUs (M3), despite increasing model complexity, not only do not lead to an improvement, but cause deterioration in space-time validation. Overall, these results suggest that causality relationships explaining the influence of climate and landscape characteristics on streamflow signatures can be constructively used for distributed model building.

The relatively good performance obtained in space–time validation suggests that the proposed approach could be used for the prediction of the streamflow in other ungauged locations within the Thur catchment. The method proposed uses data that is commonly available in many gauged catchments (e.g. meteorological data, streamflow measurements, and maps of elevation, geology, land use, and soil); therefore, it is easily transferable to other locations.

Appendix

Appendix A: Hydrological model details

A.1 Model equations

The equations of the model are listed in this appendix; the model structure in presented in Fig. 6. Table A1 contains the model parameters with the range of variability used in calibration, Table A2 lists the water–budget equations, Table A3 and A4 present the functions and the constitutive functions used.

Team list

Marco Dal Molin, Mario Schirmer, Massimiliano Zappa, Fabrizio Fenicia.

Author contribution

MDM and FF designed all the experiments. MZ contributed in the preparation of the input data for the study. MDM conducted all the experiments and analysed the results. MDM prepared the paper with the contributions from all the authors.

Competing interests

The authors declare that they have no conflict of interest.

Acknowledgements

This study was funded by the Swiss National Science Foundation (grant 200021_169003). The authors thank Federal Office of Meteorology and Climatology MeteoSwiss for the meteorological data and the Federal Office for the Environment FOEN for the streamflow data. The authors thank Dr. Lieke Melsen and the anonymous referee for their feedback and their help in improving this paper. MZ and MS dedicate this work on the memory of Dr. Joachim Gurtz, who pioneered the work on distributed hydrological modelling in the Thur catchment (Gurtz et al., 1999).

References

Abbaspour, K. C., Yang, J., Maximov, I., Siber, R., Bogner, K., Mieleitner, J., Zobrist, J., and Srinivasan, R.: Modelling hydrology and water quality in the pre-alpine/alpine Thur watershed using SWAT, J Hydrol, 333, 413-430, 10.1016/j.jhydrol.2006.09.014, 2007.

Addor, N., Newman, A. J., Mizukami, N., and Clark, M. P.: The CAMELS data set: catchment attributes and meteorology for large-sample studies, Hydrol Earth Syst Sc, 21, 5293-5313, 10.5194/hess-21-5293-2017, 2017.

Ammann, L., Reichert, P., and Fenicia, F.: A framework for likelihood functions of deterministic hydrological models, Hydrol. Earth Syst. Sci. Discuss., 2018, 1-39, 10.5194/hess-2018-406, 2018.

Andréassian, V., Perrin, C., Berthet, L., Le Moine, N., Lerat, J., Loumagne, C., Oudin, L., Mathevet, T., Ramos, M. H., and Valéry, A.: HESS Opinions "Crash tests for a standardized evaluation of hydrological models", Hydrol. Earth Syst. Sci., 13, 1757-1764, 10.5194/hess-13-1757-2009, 2009.

Antonetti, M., Buss, R., Scherrer, S., Margreth, M., and Zappa, M.: Mapping dominant runoff processes: an evaluation of different approaches using similarity measures and synthetic runoff simulations, Hydrol. Earth Syst. Sci., 20, 2929-2945, 10.5194/hess-20-2929-2016, 2016.

Artusi, R., Verderio, P., and Marubini, E.: Bravais-Pearson and Spearman Correlation Coefficients: Meaning, Test of Hypothesis and Confidence Interval, The International Journal of Biological Markers, 17, 148-151, 10.1177/172460080201700213, 2002.

Baker, D. B., Richards, R. P., Loftus, T. T., and Kramer, J. W.: A new flashiness index: Characteristics and applications to midwestern rivers and streams, J Am Water Resour As, 40, 503-522, DOI 10.1111/j.1752-1688.2004.tb01046.x, 2004.

Berger, K. P., and Entekhabi, D.: Basin hydrologic response relations to distributed physiographic descriptors and climate, J Hydrol, 247, 169-182, Doi 10.1016/S0022-1694(01)00383-3, 2001.

Bloomfield, J. P., Allen, D. J., and Griffiths, K. J.: Examining geological controls on baseflow index (BFI) using regression analysis: An illustration from the Thames Basin, UK, J Hydrol, 373, 164-176, 10.1016/j.jhydrol.2009.04.025, 2009.

Bouaziz, L., Weerts, A., Schellekens, J., Sprokkereef, E., Stam, J., Savenije, H., and Hrachowitz, M.: Redressing the balance: quantifying net intercatchment groundwater flows, Hydrol Earth Syst Sc, 22, 6415-6434, 2018.

Box, G. E. P., and Cox, D. R.: An Analysis of Transformations, J R Stat Soc B, 26, 211-252, 1964.

Brunner, M. I., Furrer, R., and Favre, A. C.: Modeling the spatial dependence of floods using the Fisher copula, Hydrol. Earth Syst. Sci., 23, 107-124, 10.5194/hess-23-107-2019, 2019.

Court, A.: Measures of streamflow timing, Journal of Geophysical Research (1896-1977), 67, 4335-4339, doi:10.1029/JZ067i011p04335, 1962.

De Lavenne, A., Thirel, G., Andréassian, V., Perrin, C., and Ramos, M.-H.: Spatial variability of the parameters of a semi-distributed hydrological model, 7th International Water Resources Management Conference of ICWRS, 2016, 87-94,

Eckhardt, K.: A comparison of baseflow indices, which were calculated with seven different baseflow separation methods, J Hydrol, 352, 168-173, 10.1016/j.jhydrol.2008.01.005, 2008.

Falkenmark, M., and Chapman, T.: Comparative hydrology: An ecological approach to land and water resources, The Unesco Press, 1989.

Fenicia, F., Kavetski, D., and Savenije, H. H. G.: Elements of a flexible approach for conceptual hydrological modeling: 1. Motivation and theoretical development, Water Resour Res, 47, Artn W11510
10.1029/2010wr010174, 2011.

Fenicia, F., Kavetski, D., Savenije, H. H. G., Clark, M. P., Schoups, G., Pfister, L., and Freer, J.: Catchment properties, function, and conceptual model representation: is there a correspondence?, Hydrol Process, 28, 2451-2467, 10.1002/hyp.9726, 2014.

Fenicia, F., Kavetski, D., Savenije, H. H. G., and Pfister, L.: From spatially variable streamflow to distributed hydrological models: Analysis of key modeling decisions, Water Resour Res, 52, 954-989, 10.1002/2015wr017398, 2016.

Feyen, L., Kalas, M., and Vrugt, J. A.: Semi-distributed parameter optimization and uncertainty assessment for large-scale streamflow simulation using global optimization/Optimisation de paramètres semi-distribués et évaluation de l'incertitude pour la simulation de débits à grande échelle par l'utilisation d'une optimisation globale, Hydrological Sciences Journal, 53, 293-308, 2008.

Fundel, F., Jorg-Hess, S., and Zappa, M.: Monthly hydrometeorological ensemble prediction of streamflow droughts and corresponding drought indices, Hydrol Earth Syst Sc, 17, 395-407, 10.5194/hess-17-395-2013, 2013.

Garen, D. C., and Marks, D.: Spatial fields of meteorological input data including forest canopy corrections for an energy budget snow simulation model, IAHS PUBLICATION, 349-354, 2001.

Gurtz, J., Baltensweiler, A., and Lang, H.: Spatially distributed hydrotope-based modelling of evapotranspiration and runoff in mountainous basins, Hydrol Process, 13, 2751-2768, Doi 10.1002/(Sici)1099-1085(19991215)13:17<2751::Aid-Hyp897>3.3.Co;2-F, 1999.

Gurtz, J., Verbunt, M., Zappa, M., Moesch, M., Pos, F., and Moser, U.: Long-term hydrometeorological measurements and model-based analyses in the hydrological research catchment Rietholzbach, Journal of Hydrology and Hydromechanics, 51, 162-174, 2003.

Hrachowitz, M., Savenije, H., Blöschl, G., McDonnell, J., Sivapalan, M., Pomeroy, J., Arheimer, B., Blume, T., Clark, M., and Ehret, U.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrological sciences journal, 58, 1198-1255, 2013.

Hurford, A. P., and Harou, J. J.: Balancing ecosystem services with energy and food security - Assessing trade-offs from reservoir operation and irrigation investments in Kenya's Tana Basin, Hydrol Earth Syst Sc, 18, 3259-3277, 10.5194/hess-18-3259-2014, 2014.

Jasper, K., Calanca, P., Gyalistras, D., and Fuhrer, J.: Differential impacts of climate change on the hydrology of two alpine river basins, Climate Res, 26, 113-129, DOI 10.3354/cr026113, 2004.

Jorg-Hess, S., Kempf, S. B., Fundel, F., and Zappa, M.: The benefit of climatological and calibrated reforecast data for simulating hydrological droughts in Switzerland, Meteorol Appl, 22, 444-458, 10.1002/met.1474, 2015.

Kavetski, D., and Kuczera, G.: Model smoothing strategies to remove microscale discontinuities and spurious secondary optima in objective functions in hydrological calibration, Water Resour Res, 43, Artn W03411
10.1029/2006wr005195, 2007.

Kavetski, D., Kuczera, G., Thyer, M., and Renard, B.: Multistart Newton-type optimisation methods for the calibration of conceptual hydrological models, Modsim 2007: International Congress on Modelling and Simulation, 2513-2519, 2007.

Kavetski, D., Fenicia, F., and Clark, M. P.: Impact of temporal data resolution on parameter inference and model identification in conceptual hydrological modeling: Insights from an experimental catchment, Water Resour Res, 47, Artn W05501
10.1029/2010wr009525, 2011.

Kirchner, J. W., Feng, X. H., Neal, C., and Robson, A. J.: The fine structure of water-quality dynamics: the (high-frequency) wave of the future, Hydrol Process, 18, 1353-1359, 10.1002/hyp.5537, 2004.

Kroll, C. N., and Song, P.: Impact of multicollinearity on small sample hydrologic regression models, Water Resour Res, 49, 3756-3769, 10.1002/wrcr.20315, 2013.

Kuentz, A., Arheimer, B., Hundecha, Y., and Wagener, T.: Understanding hydrologic variability across Europe through catchment classification, Hydrol Earth Syst Sc, 21, 2863-2879, 10.5194/hess-21-2863-2017, 2017.

Lacey, G. C., and Grayson, R. B.: Relating baseflow to catchment properties in south-eastern Australia, J Hydrol, 204, 231-250, Doi 10.1016/S0022-1694(97)00124-8, 1998.

Ladson, A. R., Brown, R., Neal, B., and Nathan, R.: A Standard Approach to Baseflow Separation Using The Lyne and Hollick Filter, Australasian Journal of Water Resources, 17, 25-34, 10.7158/13241583.2013.11465417, 2013.

Lerat, J., Andreassian, V., Perrin, C., Vaze, J., Perraud, J.-M., Ribstein, P., and Loumagne, C.: Do internal flow measurements improve the calibration of rainfall-runoff models?, Water Resour Res, 48, 2012.

Lindstrom, G., Johansson, B., Persson, M., Gardelin, M., and Bergstrom, S.: Development and test of the distributed HBV-96 hydrological model, J Hydrol, 201, 272-288, Doi 10.1016/S0022-1694(97)00041-3, 1997.

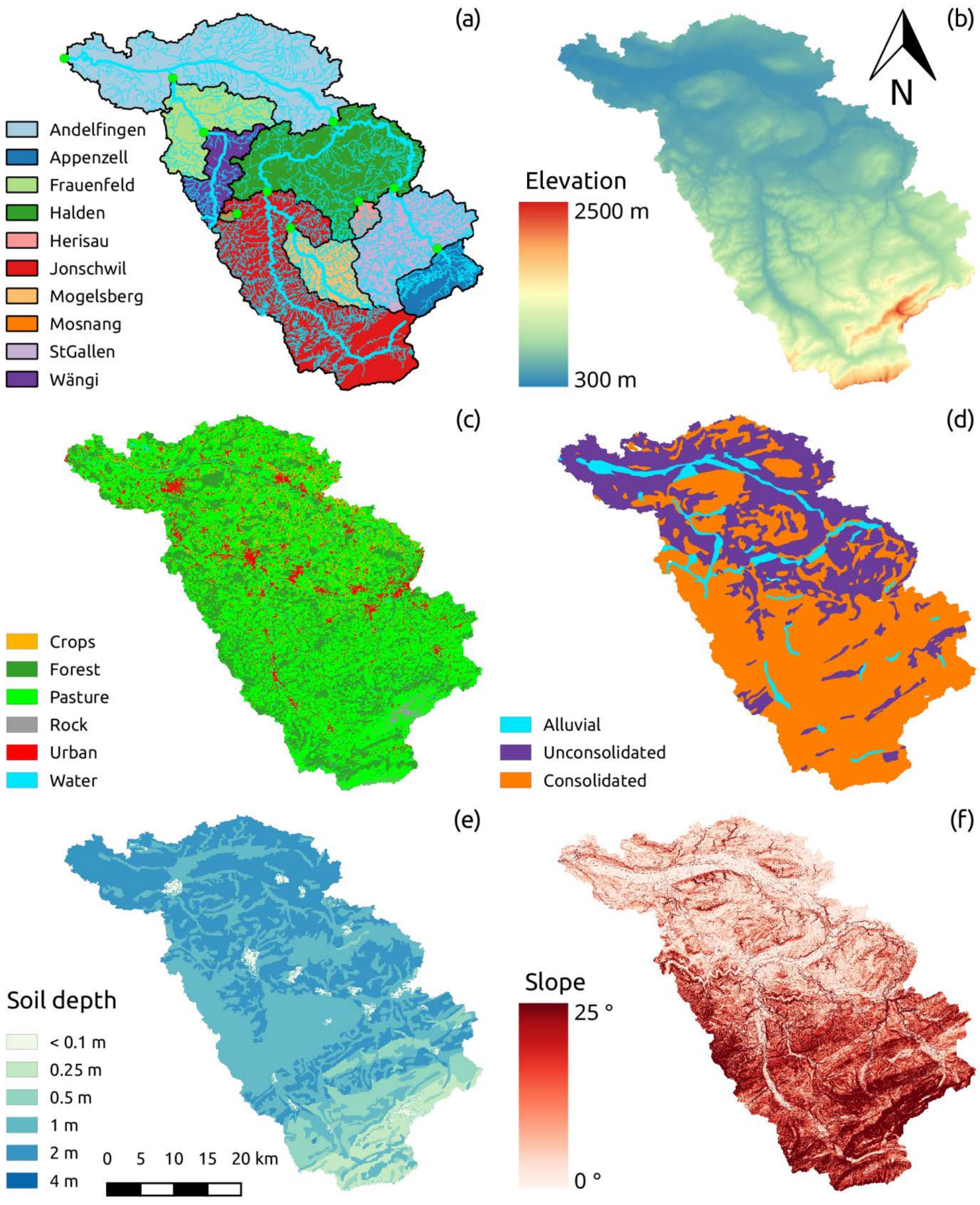
Mazvimavi, D., Meijerink, A. M. J., Savenije, H. H. G., and Stein, A.: Prediction of flow characteristics using multiple regression and neural networks: A case study in Zimbabwe, Phys Chem Earth, 30, 639-647, 10.1016/j.pce.2005.08.003, 2005.

McDonnell, J. J., and Woods, R.: On the need for catchment classification, J Hydrol, 299, 2-3, 2004.

McInerney, D., Thyer, M., Kavetski, D., Lerat, J., and Kuczera, G.: Improving probabilistic prediction of daily streamflow by identifying Pareto optimal approaches for modeling heteroscedastic residual errors, Water Resour Res, 53, 2199-2239, 10.1002/2016wr019168, 2017.

Melsen, L., Teuling, A., Torfs, P., Zappa, M., Mizukami, N., Clark, M., and Uijlenhoet, R.: Representation of spatial and temporal variability in large-domain hydrological models: case study for a mesoscale pre-Alpine basin, Hydrol Earth Syst Sc, 20, 2207-2226, 10.5194/hess-20-2207-2016, 2016.
Melsen, L. A., Teuling, A. J., van Berkum, S. W., Torfs, P. J. J. F., and Uijlenhoet, R.: Catchments as simple dynamical systems: A case study on methods and data requirements for parameter identification, Water Resour Res, 50, 5577-5596, doi:10.1002/2013WR014720, 2014.
Menzel, L.: Modelling canopy resistances and transpiration of grassland, Phys Chem Earth, 21, 123-129, https://doi.org/10.1016/S0079-1946(97)85572-3, 1996.
Miller, A., Tibshirani, R., Cox, D., Keiding, N., Isham, V., Louis, T., Tong, H., and Reid, N.: Subset Selection in Regression, Chapman and Hall/CRC New York 2002.
Monteith, J. L.: Vegetation and the atmosphere, Academic Press, London ; New York, 1975.
Parajka, J., Merz, R., and Blöschl, G.: A comparison of regionalisation methods for catchment model parameters, Hydrol. Earth Syst. Sci., 9, 157-171, 10.5194/hess-9-157-2005, 2005.
Sawicz, K., Wagener, T., Sivapalan, M., Troch, P. A., and Carrillo, G.: Catchment classification: empirical analysis of hydrologic similarity based on catchment function in the eastern USA, Hydrol. Earth Syst. Sci., 15, 2895-2911, 10.5194/hess-15-2895-2011, 2011.
Scherrer, S., and Naef, F.: A decision scheme to indicate dominant hydrological flow processes on temperate grassland, Hydrol Process, 17, 391-401, 10.1002/hyp.1131, 2003.
Schirmer, M., Luster, J., Linde, N., Perona, P., Mitchell, E. A. D., Barry, D. A., Hollender, J., Cirpka, O. A., Schneider, P., Vogt, T., Radny, D., and Durisch-Kaiser, E.: Morphological, hydrological, biogeochemical and ecological changes and challenges in river restoration – the Thur River case study, Hydrol. Earth Syst. Sci., 18, 2449-2462, 10.5194/hess-18-2449-2014, 2014.
Seneviratne, S. I., Lehner, I., Gurtz, J., Teuling, A. J., Lang, H., Moser, U., Grebner, D., Menzel, L., Schroff, K., Vitvar, T., and Zappa, M.: Swiss prealpine Rietholzbach research catchment and lysimeter: 32 year time series and 2003 drought event, Water Resour Res, 48, Artn W06526
10.1029/2011wr011749, 2012.
Sivapalan, M.: Pattern, Process and Function: Elements of a Unified Theory of Hydrology at the Catchment Scale, in: Encyclopedia of Hydrological Sciences, 2006.
Toth, E.: Catchment classification based on characterisation of streamflow and precipitation time series, Hydrol Earth Syst Sc, 17, 1149-1159, 10.5194/hess-17-1149-2013, 2013.
Trancoso, R., Phinn, S., McVicar, T. R., Larsen, J. R., and McAlpine, C. A.: Regional variation in streamflow drivers across a continental climatic gradient, Ecohydrology, 10, UNSP e1816
10.1002/eco.1816, 2017.
van Esse, W. R., Perrin, C., Booij, M. J., Augustijn, D. C. M., Fenicia, F., Kavetski, D., and Lobligeois, F.: The influence of conceptual model structure on model performance: a comparative study for 237 French catchments, Hydrol Earth Syst Sc, 17, 4227-4239, 10.5194/hess-17-4227-2013, 2013.
Verbunt, M., Zappa, M., Gurtz, J., and Kaufmann, P.: Verification of a coupled hydrometeorological modelling approach for alpine tributaries in the Rhine basin, J Hydrol, 324, 224-238, https://doi.org/10.1016/j.jhydrol.2005.09.036, 2006.
Viviroli, D., Zappa, M., Gurtz, J., and Weingartner, R.: An introduction to the hydrological modelling system PREVAH and its pre- and post-processing-tools, Environ Modell Softw, 24, 1209-1222, 10.1016/j.envsoft.2009.04.001, 2009.
von Freyberg, J., Radny, D., Gall, H. E., and Schirmer, M.: Implications of hydrologic connectivity between hillslopes and riparian zones on streamflow composition, J Contam Hydrol, 169, 62-74, 10.1016/j.jconhyd.2014.07.005, 2014.
von Freyberg, J., Moeck, C., and Schirmer, M.: Estimation of groundwater recharge and drought severity with varying model complexity, J Hydrol, 527, 844-857, 10.1016/j.jhydrol.2015.05.025, 2015.
Wagener, T., Sivapalan, M., McDonnell, J., Hooper, R., Lakshmi, V., Liang, X., and Kumar, P.: Predictions in ungauged basins as a catalyst for multidisciplinary hydrology, Eos, Transactions American Geophysical Union, 85, 451-457, 2004.
Wagener, T., Sivapalan, M., Troch, P., and Woods, R.: Catchment Classification and Hydrologic Similarity, Geography Compass, 1, 901-931, doi:10.1111/j.1749-8198.2007.00039.x, 2007.

Figures


(a)
Andelfingen
Appenzell
Frauenfeld
Halden
Herisau
Jonschwil
Mogelsberg
Mosnang
StGallen
Wängi
(b)
N
Elevation
2500 m
300 m
(c)
Crops
Forest
Pasture
Rock
Urban
Water
(d)
Alluvial
Unconsolidated
Consolidated
(e)
Soil depth
< 0.1 m
0.25 m
0.5 m
1 m
2 m
4 m
0 5 10 15 20 km
(f)
Slope
25 °
0 °

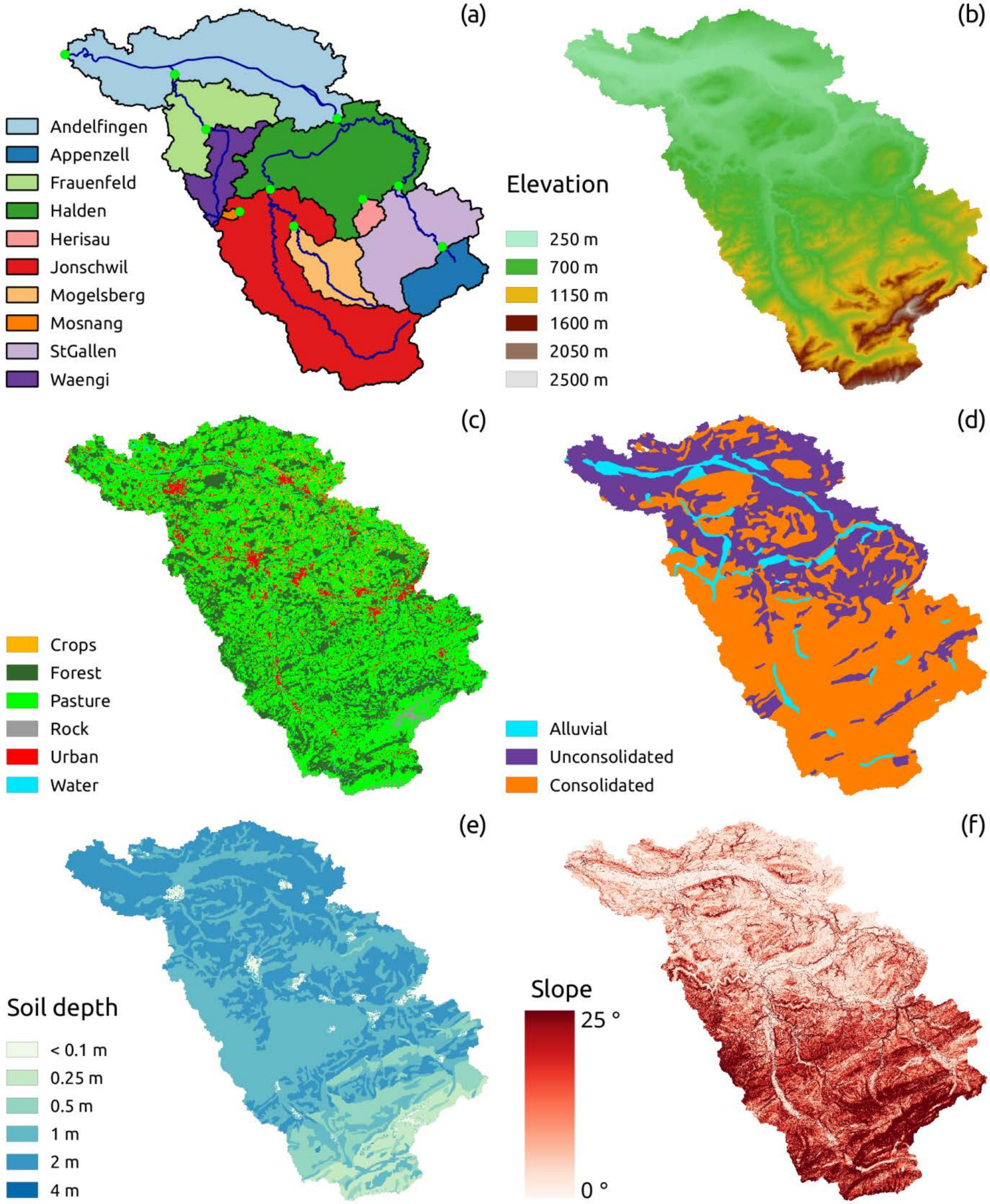
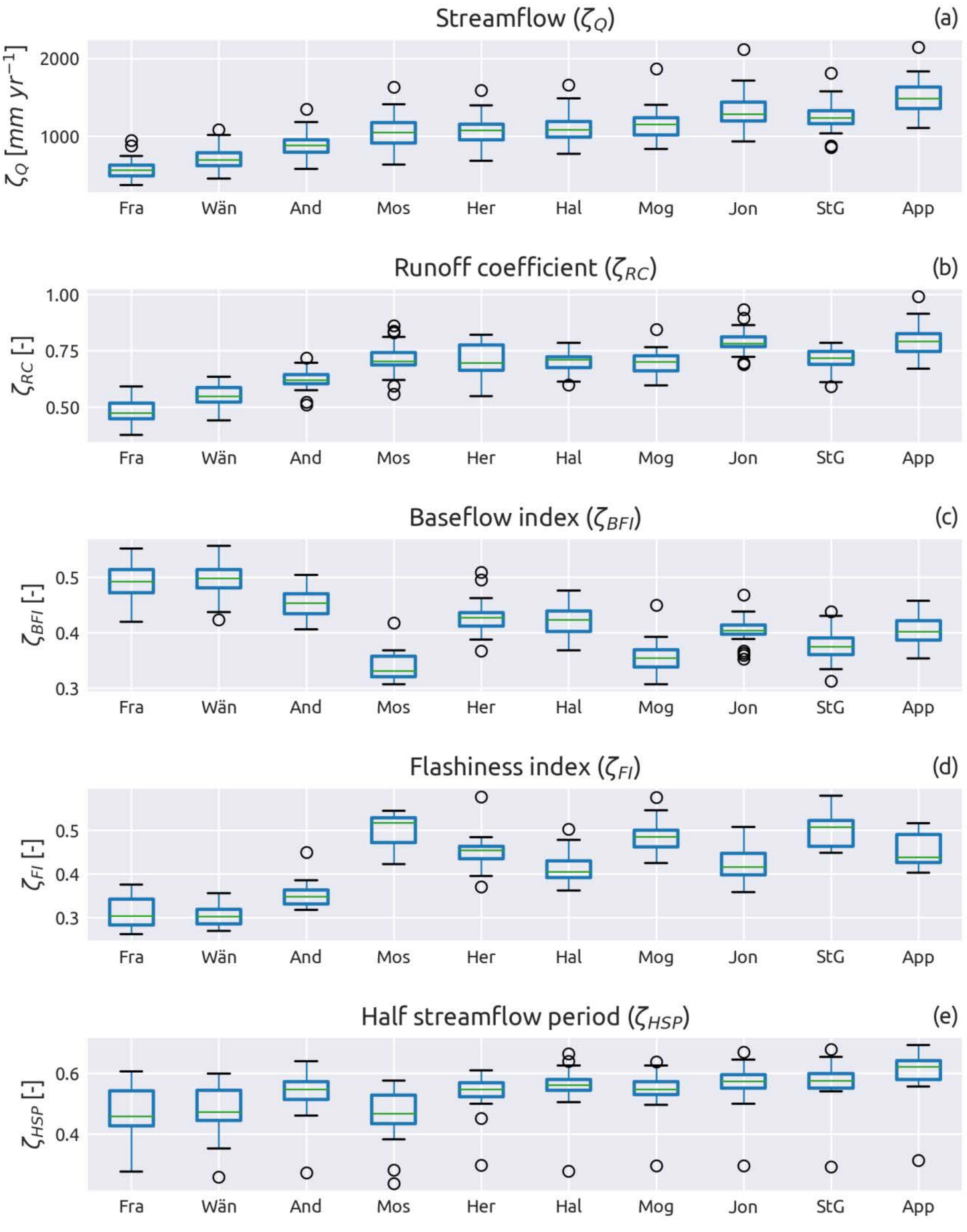
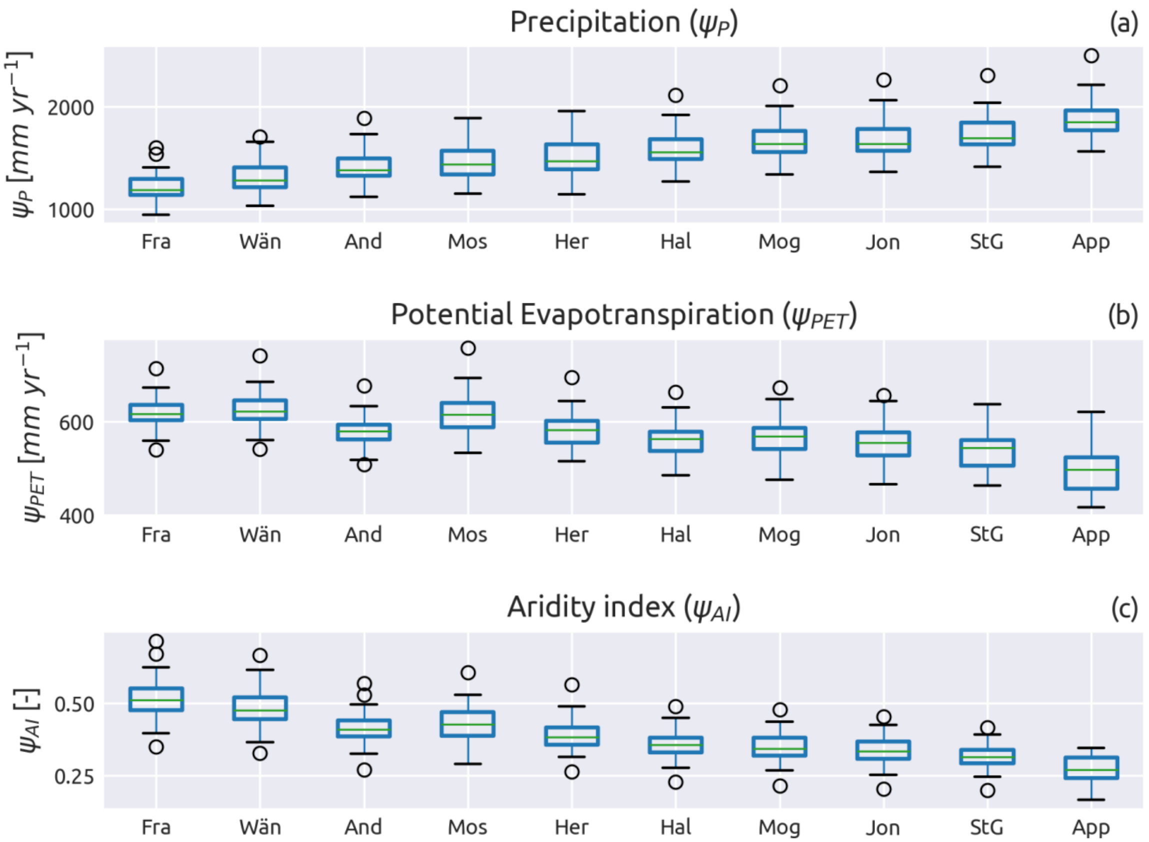

(a)
Andelfingen
Appenzell
Frauenfeld
Halden
Herisau
Jonschwil
Mogelsberg
Mosnang
StGallen
Waengi
(b)
Elevation
250 m
700 m
1150 m
1600 m
2050 m
2500 m
(c)
Crops
Forest
Pasture
Rock
Urban
Water
(d)
Alluvial
Unconsolidated
Consolidated
(e)
Soil depth
< 0.1 m
0.25 m
0.5 m
1 m
2 m
4 m
(f)
Slope
25 °
0 °

Figure 1: Landscape characteristics of the Thur catchment: (a) subdivision in subcatchments, river network, and gauging stations; (b) elevation map; (c) land use map; (d) simplified geology map; (e) soil depth map; (f) slope map (derived from the elevation map).

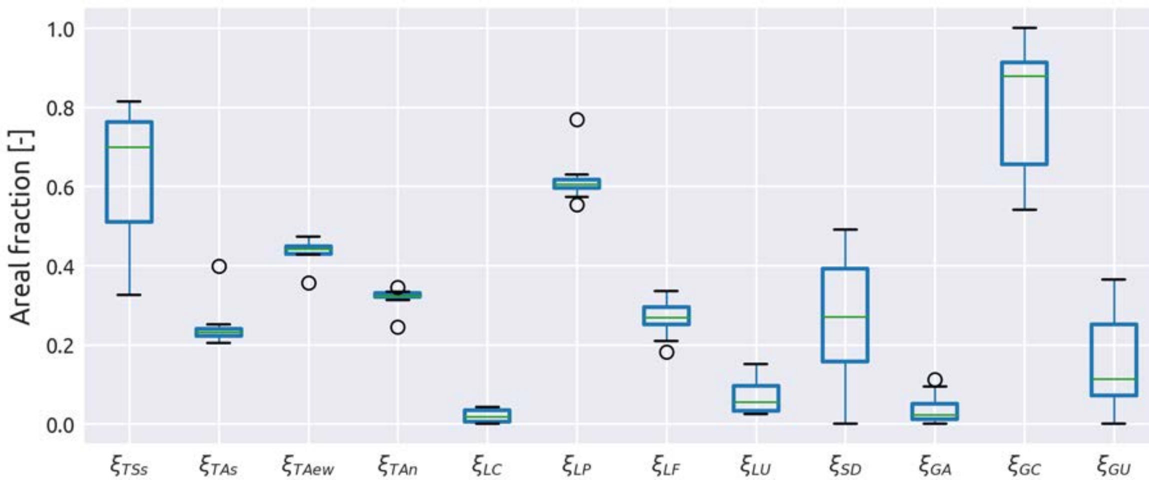


Streamflow (ζ_Q)
(a)
ζ_Q [$mm\ yr^{-1}$]
2000
1000
Fra
Wän
And
Mos
Her
Hal
Mog
Jon
StG
App
Runoff coefficient (ζ_{RC})
(b)
ζ_{RC} [-]
1.00
0.75
0.50
Baseflow index (ζ_{BFI})
(c)
ζ_{BFI} [-]
0.5
0.4
0.3
Flashiness index (ζ_{FI})
(d)
ζ_{FI} [-]
0.5
0.4
0.3
Half streamflow period (ζ_{HSP})
(e)
ζ_{HSP} [-]
0.6
0.4

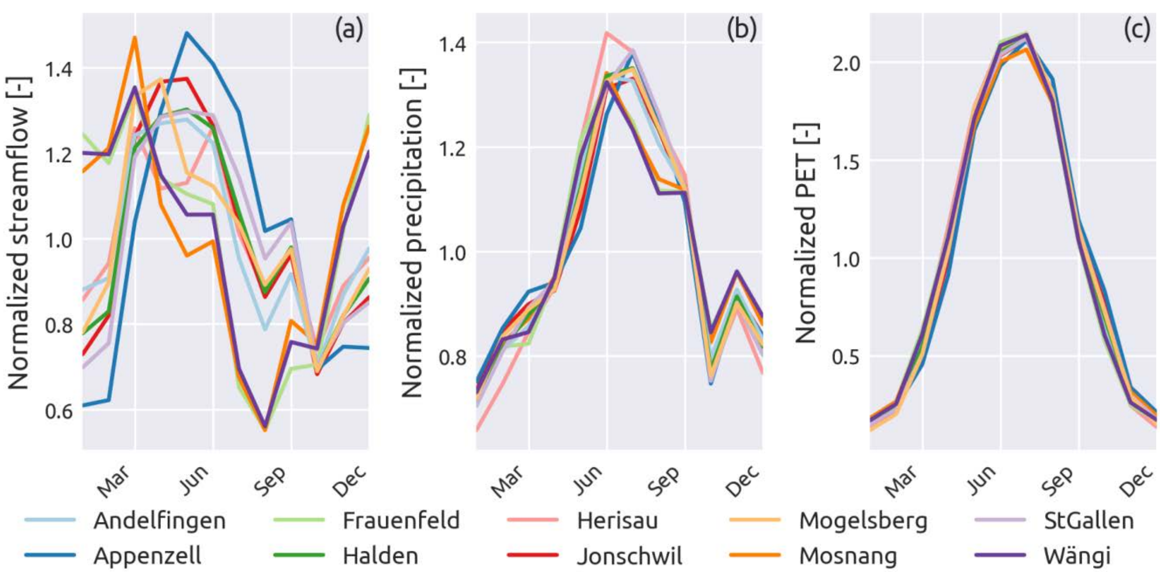
~~**Figure 2: Interannual variability of the streamflow signatures of the subcatchments; the catchments are sorted according to their mean elevation and the names are abbreviated using the first three letters.**~~



~~**Figure 3: Interannual variability of the meteorological inputs of the subcatchments.**~~



~~**Figure 4: Variability of the landscape characteristics between the subcatchments. Characteristics that are not presented in this figure are reported in Table 1**~~



~~**Figure 5: Monthly variability of streamflow (a), precipitation (b) and potential evapotranspiration (c). The values are normalized by the annual average value of the variable for each subcatchment since here the objective is to analyse the monthly distribution and not the actual magnitude.**~~

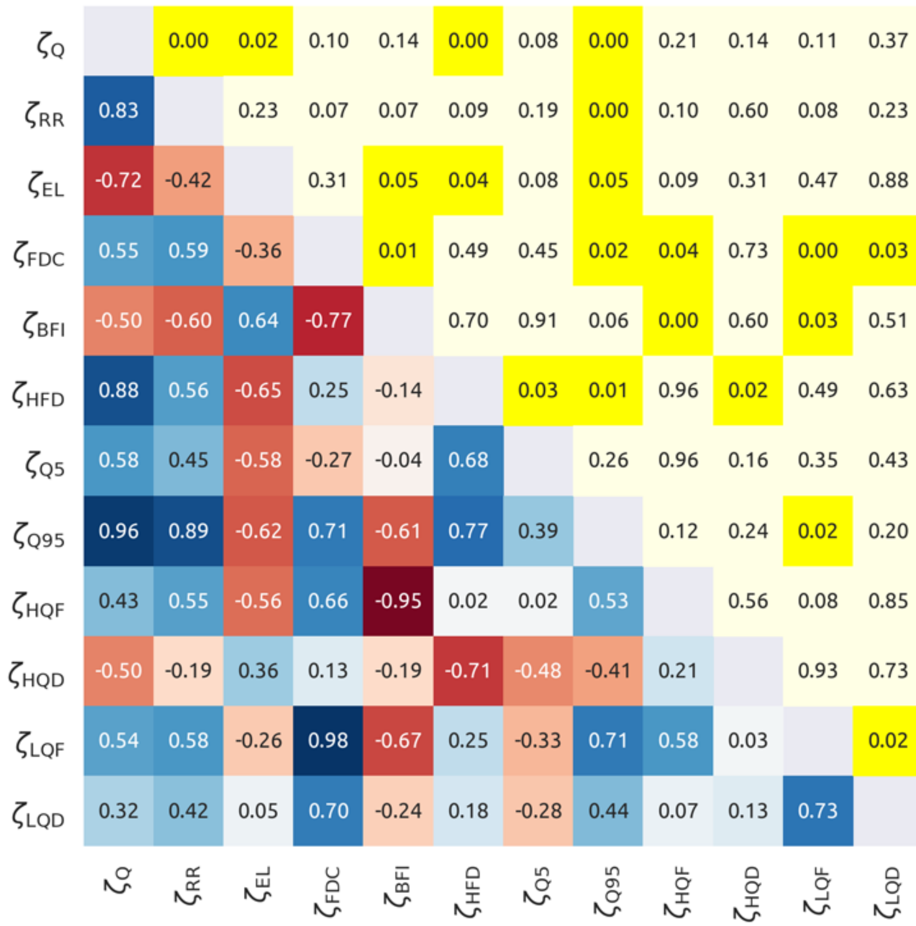


Figure 2: Internal correlation between the streamflow signatures. The lower triangle shows the Spearman's rank score with the red colour that indicates negative correlations and the blue that indicates positive correlations. The upper triangle reports the corresponding p-values, where yellow colour indicates a statistically significant correlation (p-value < 0.05). The symbols used in the figure are reported in Sect. 3.1.1.

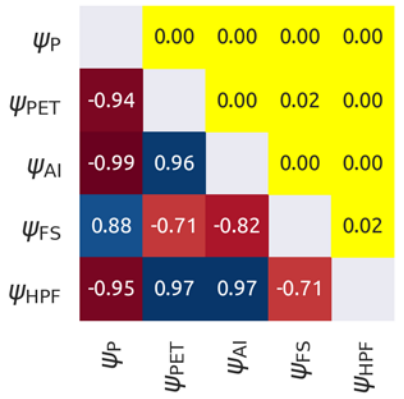


Figure 3: Internal correlation between the climatic indices. The lower triangle shows the Spearman's rank score with the red colour that indicates negative correlations and the blue that indicates positive correlations. The upper triangle reports the corresponding p-values, where yellow colour indicates a statistically significant correlation (p-value < 0.05). The symbols used in the figure are reported in Sect. 3.1.1.

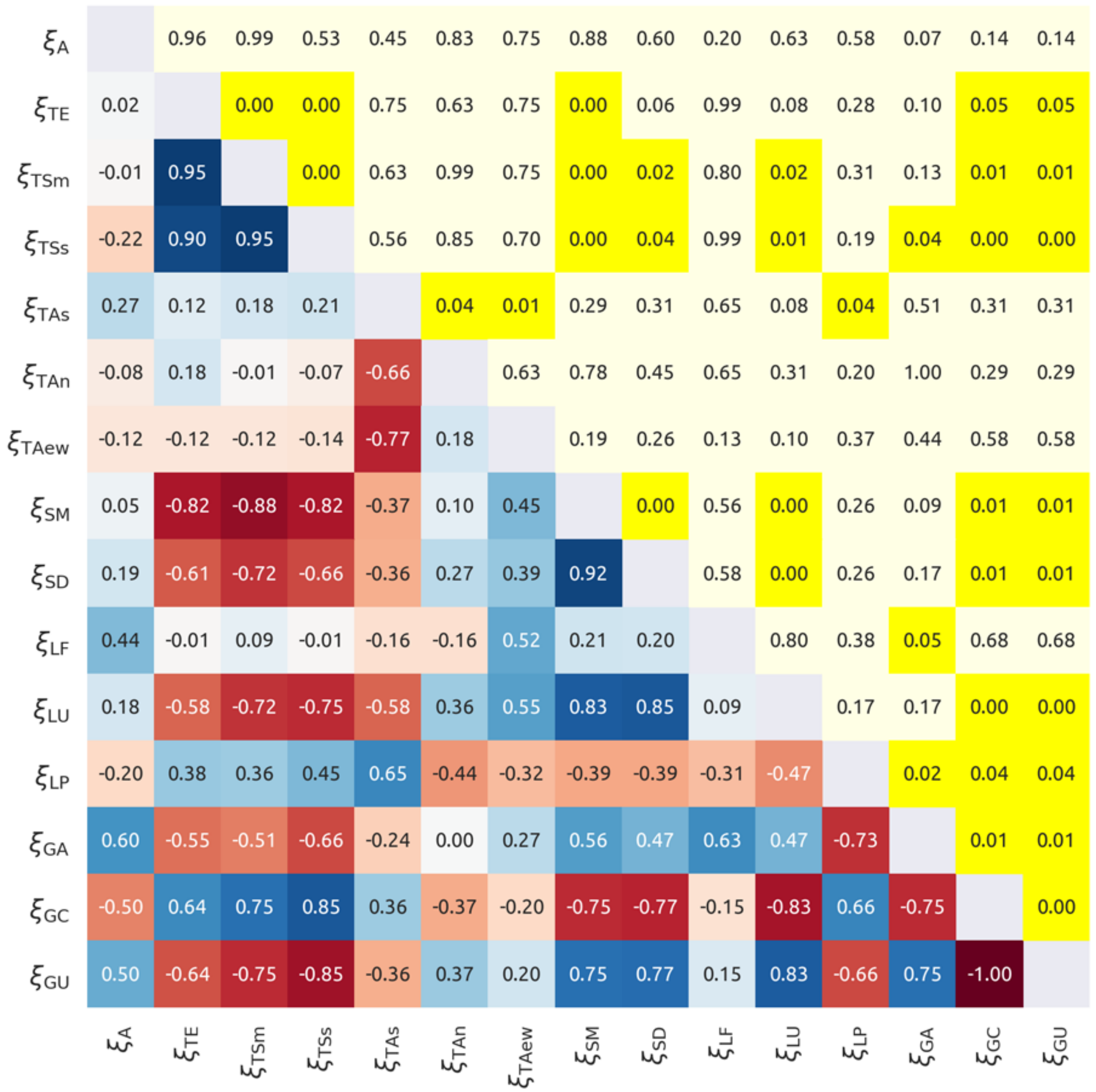


Figure 4: Internal correlation between the catchment characteristics. The lower triangle shows the Spearman's rank score with the red colour that indicates negative correlations and the blue that indicates positive correlations. The upper triangle reports the corresponding p-values, where yellow colour indicates a statistically significant correlation (p-value < 0.05). The symbols used in the figure are reported in Sect. 3.1.1.

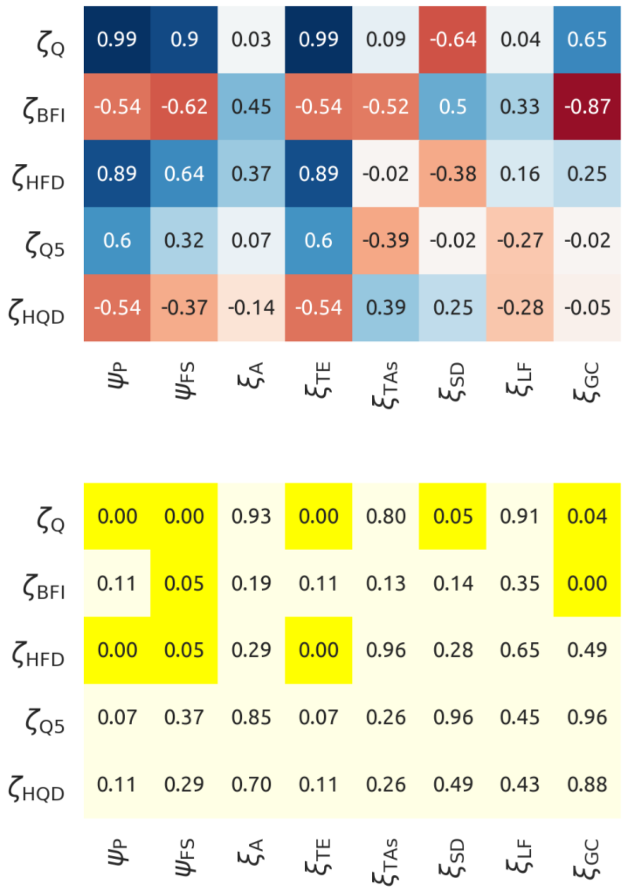


Figure 5: Correlation between the selected streamflow signatures (rows) and the selected climatic indices and catchment characteristics (columns). The upper panel shows the Spearman's rank score with the red colour that indicates negative correlations and the blue that indicates positive correlations. The lower panel reports the corresponding p-values, where yellow colour indicates a statistically significant correlation (p-value < 0.05). The symbols used in the figure are reported in Sect. 3.1.1.

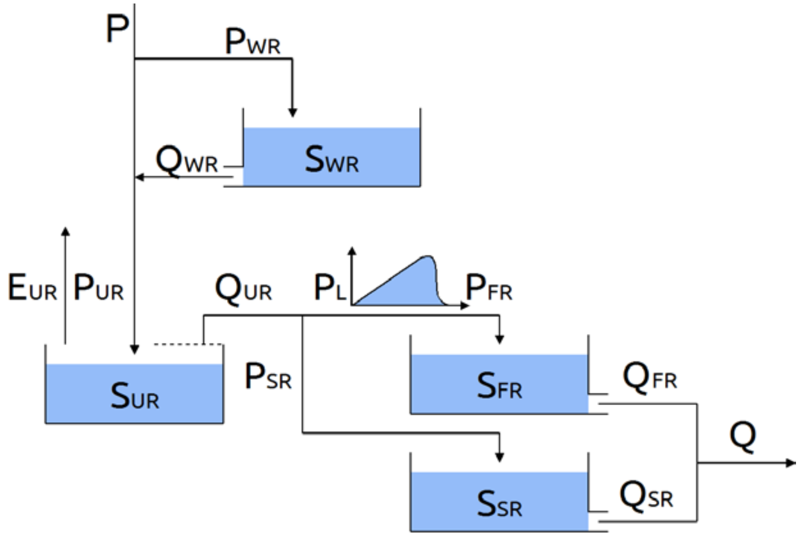
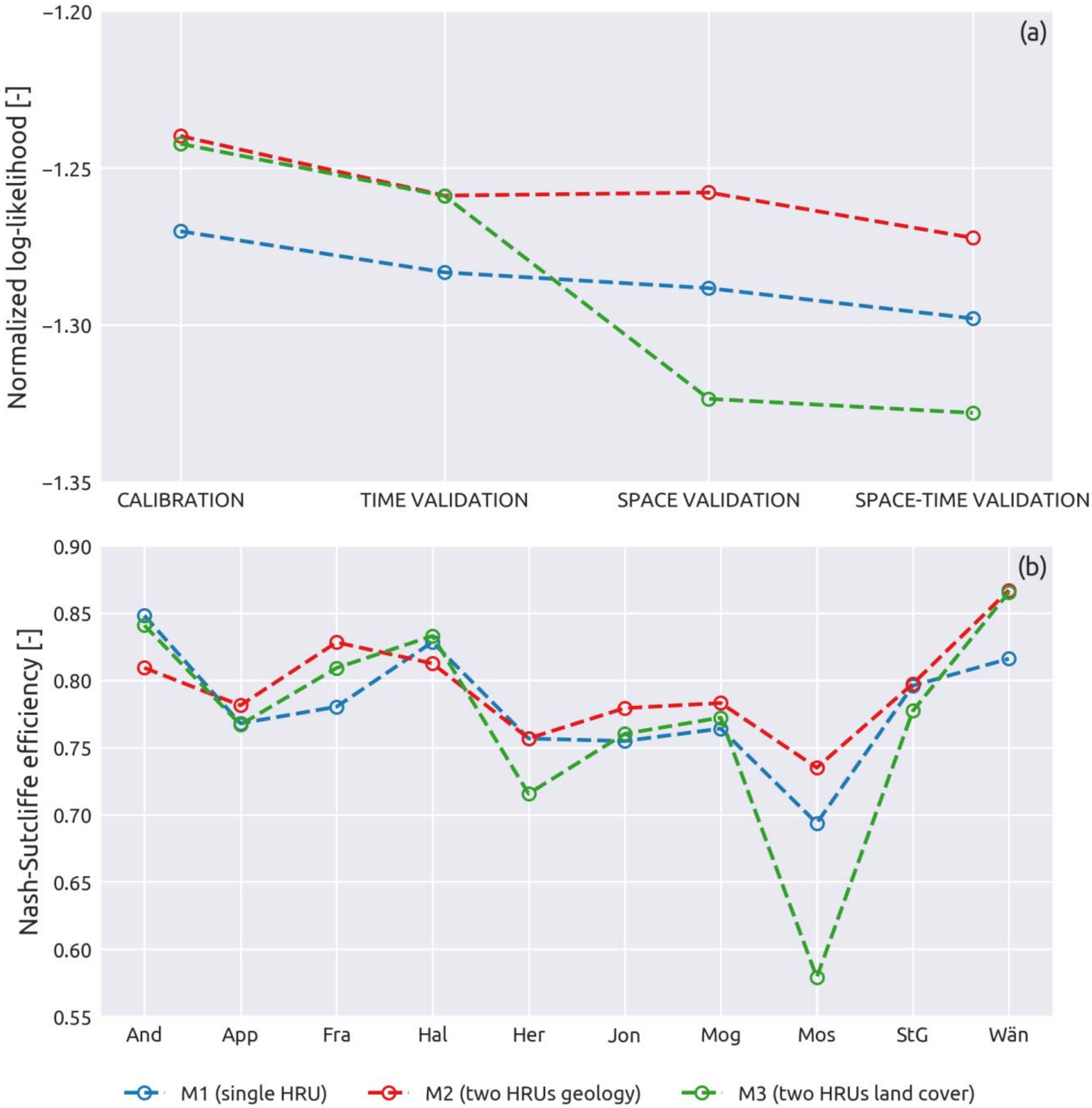


Figure 6: Schematic representation of the model structure used for the HRUs in all the model configurations. ~~The symbols and the~~In the scheme “P” represent the precipitation entering in the reservoirs, “E” the evaporation, and “Q” the outflow from the reservoirs. The subscripts indicate the reservoirs: WR = snow reservoir, UR = unsaturated reservoir, FR = fast reservoir, SR = slow reservoir, L = lag function. The governing equations are reported in Appendix A



(a)
Normalized log-likelihood [-]
−1.20
−1.25
−1.30
−1.35
CALIBRATION
TIME VALIDATION
SPACE VALIDATION
SPACE-TIME VALIDATION
(b)
Nash-Sutcliffe efficiency [-]
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
And
App
Fra
Hal
Her
Jon
Mog
Mos
StG
Wän
M1 (single HRU)
M2 (two HRUs geology)
M3 (two HRUs land cover)

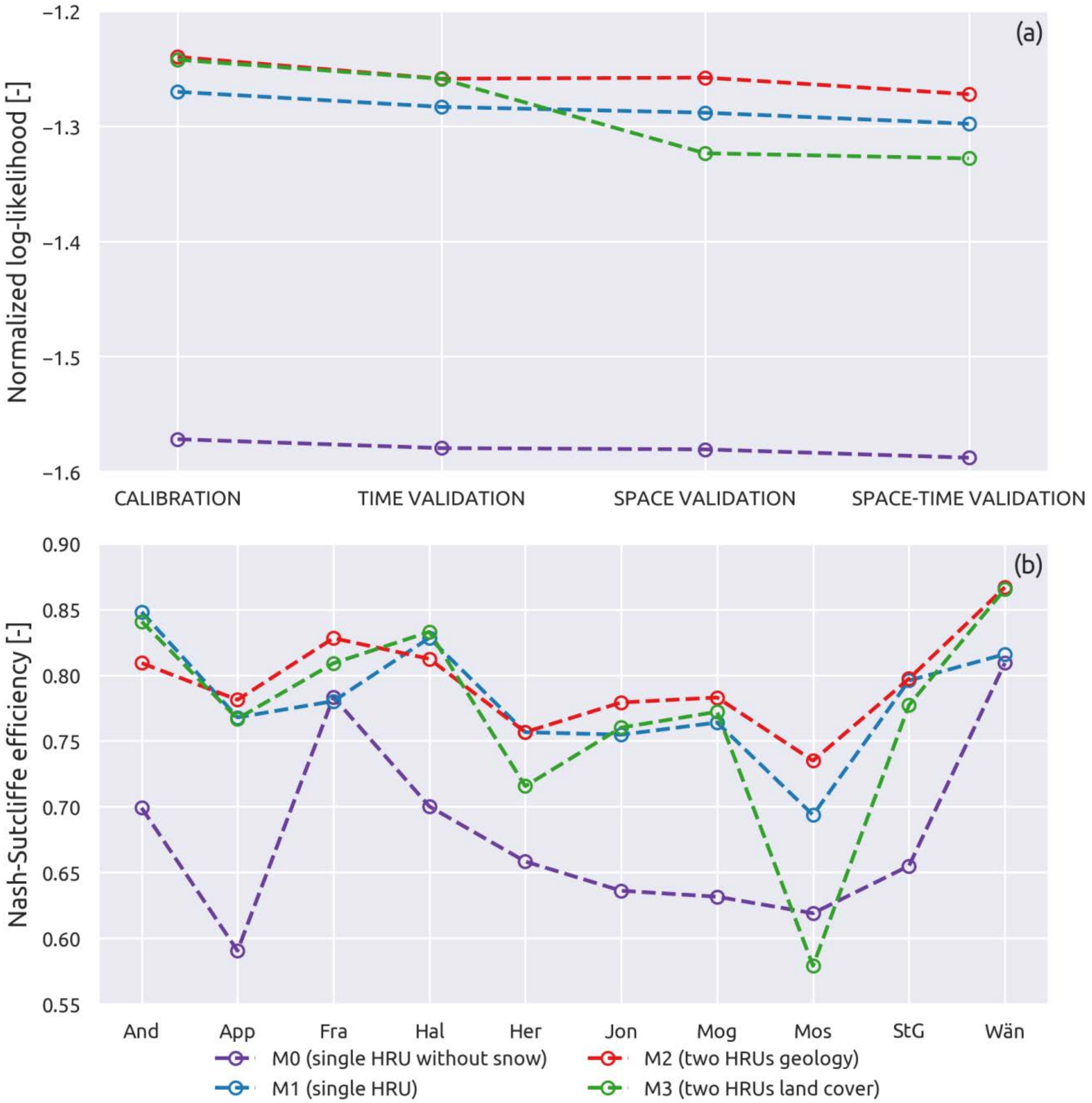
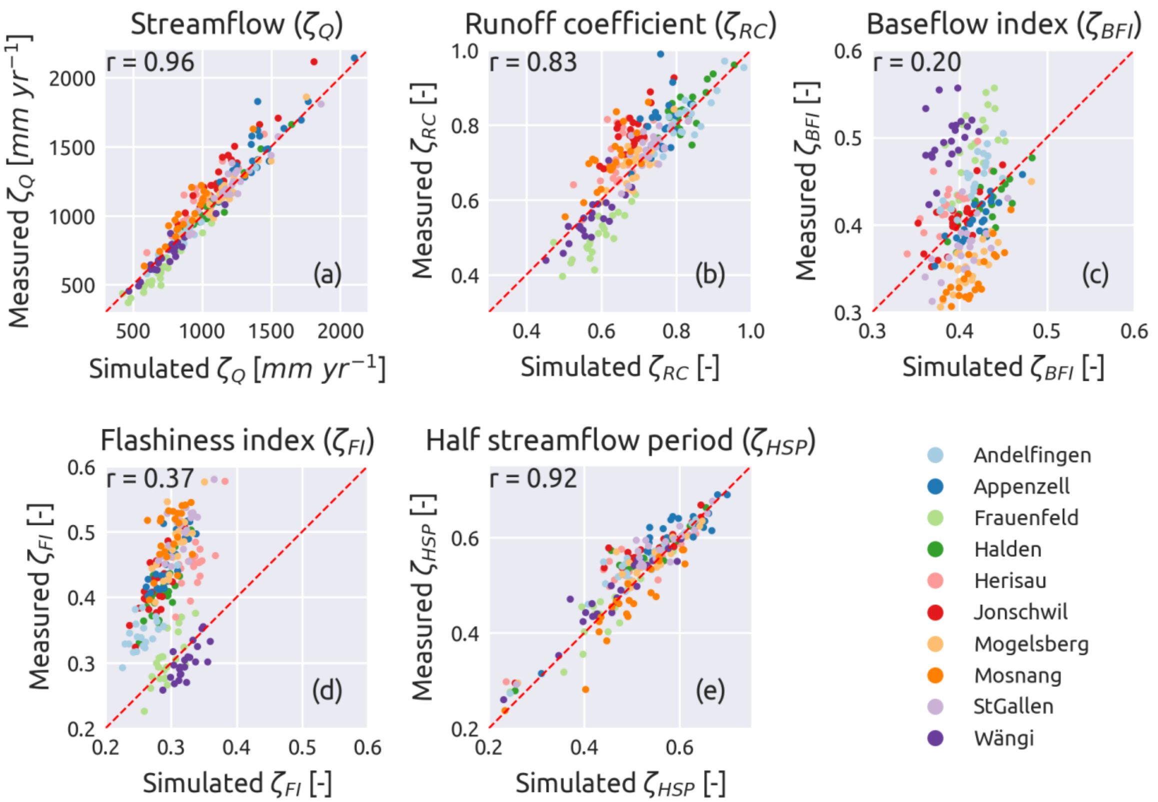


Figure 7: Normalized log–likelihood (a) and Nash–Sutcliffe efficiency (b) for the three model configurations. The upper plot (a) reports the variation between calibration and validation of the average of the 10 subcatchments; the lower plot (b) shows the variation between subcatchments during space–time validation.



Streamflow (ζ_Q)
r = 0.96
Measured ζ_Q [$mm\ yr^{-1}$]
Simulated ζ_Q [$mm\ yr^{-1}$]
(a)
Runoff coefficient (ζ_{RC})
r = 0.83
Measured ζ_{RC} [-]
Simulated ζ_{RC} [-]
(b)
Baseflow index (ζ_{BFI})
r = 0.20
Measured ζ_{BFI} [-]
Simulated ζ_{BFI} [-]
(c)
Flashiness index (ζ_{FI})
r = 0.37
Measured ζ_{FI} [-]
Simulated ζ_{FI} [-]
(d)
Half streamflow period (ζ_{HSP})
r = 0.92
Measured ζ_{HSP} [-]
Simulated ζ_{HSP} [-]
(e)
Andelfingen
Appenzell
Frauenfeld
Halden
Herisau
Jonschwil
Mogelsberg
Mosnang
StGallen
Wängi

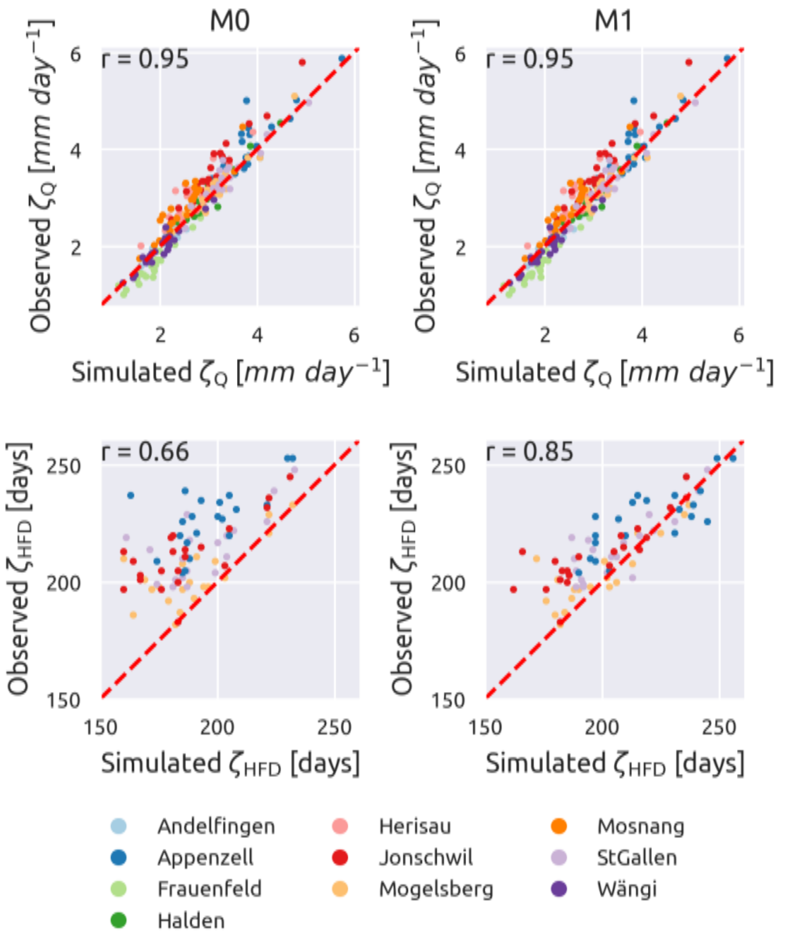
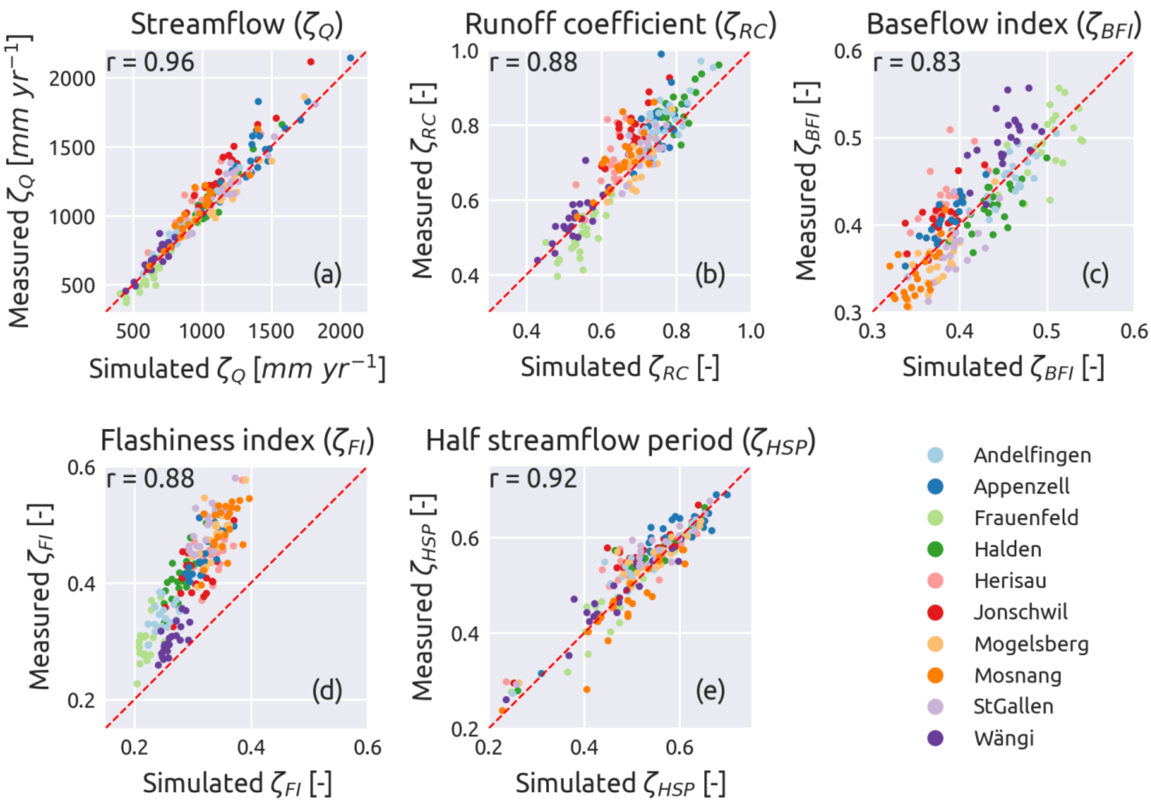
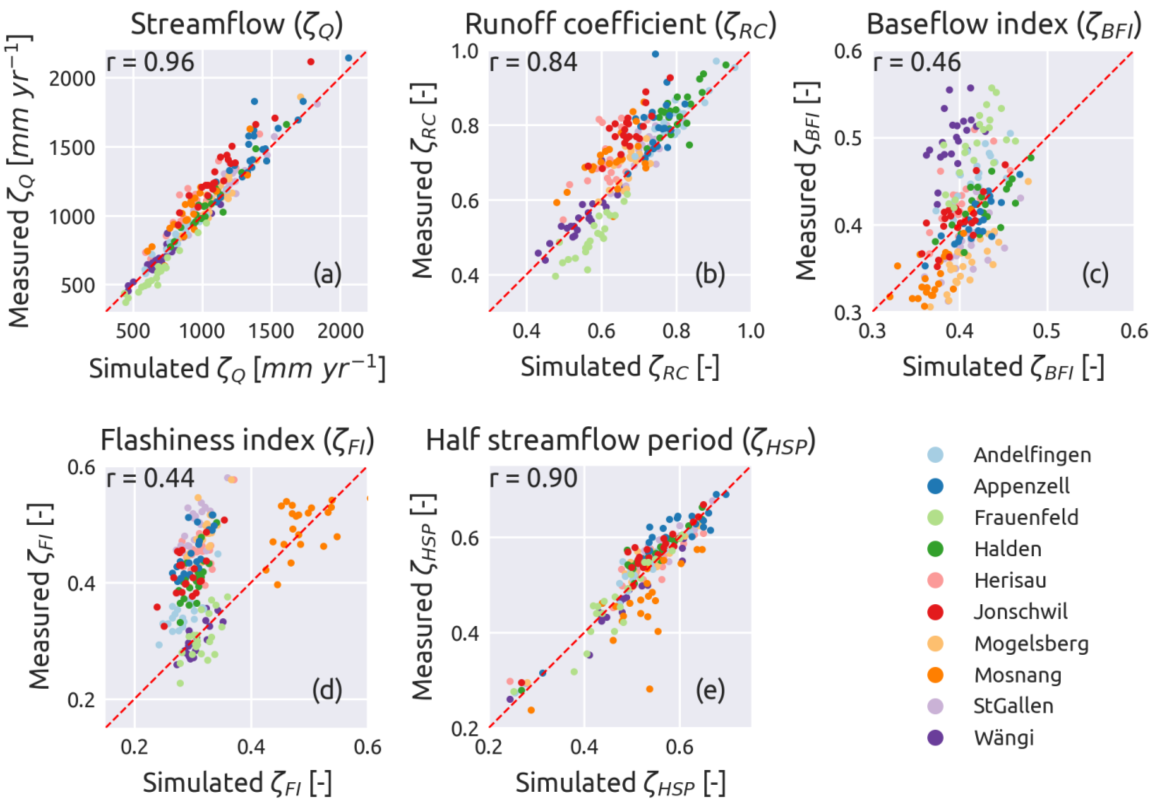


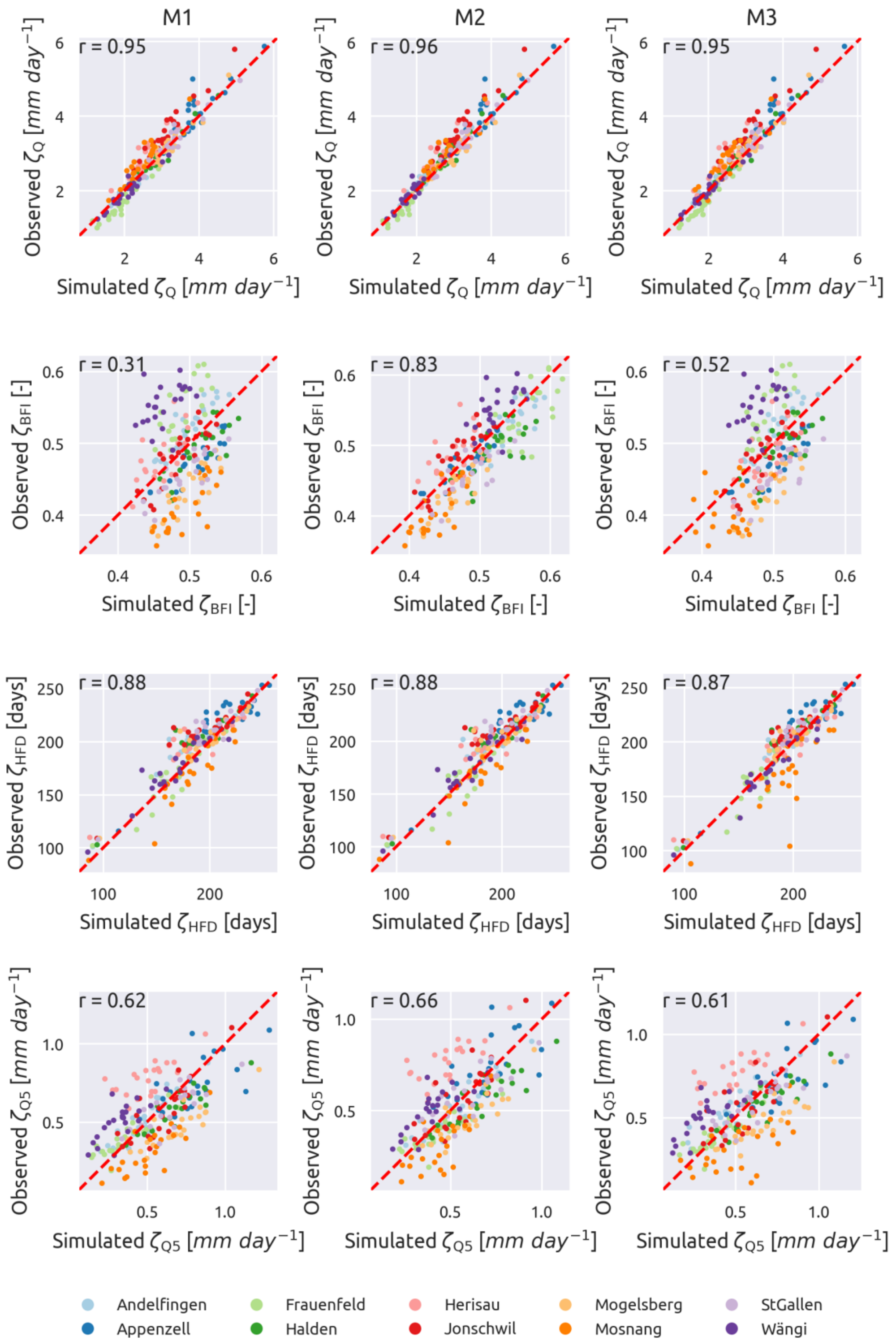
Figure 8: ~~**Single HRU model. Comparison between observed and simulated signatures in space-time validation (**~~**Influence of the model structure on the representation of the average streamflow (**~~**a), runoff coefficient (b), baseflow index (c), flashiness index (d), and**~~**ζ_Q) and the mean half streamflow** ~~**period (e)).**~~**day (ζ_{HFD}). Single HRU model without snow reservoir on the left, single HRU model with snow reservoir on the right. Each dot represents a year and each colour a subcatchment. For ζ_{HFD}, only the four subcatchments with the fraction of snow (ψ_{FS}) larger than 10 % are plotted. The red dashed line has a 45 ° slope and** ~~**represents the line**~~ **indicates where all** ~~**the**~~ **points should align in case of perfect match**~~**between simulated and observed signatures.**~~**. The** ~~**Pearson correlation coefficient**~~**Spearman's rank score (r) is also reported.**



~~Figure 9: Two HRUs model based on geology. Comparison between observed and simulated signatures in space-time validation (see caption of Fig. 8).~~



~~**Figure 10: Two HRUs model based on land use. Comparison between observed and simulated signatures in space-time validation (see caption of Fig. 8).**~~



M1
M2
M3
r = 0.95
r = 0.96
r = 0.95
Observed ζQ [mm day−1]
Simulated ζQ [mm day−1]
r = 0.31
r = 0.83
r = 0.52
Observed ζBFI [-]
Simulated ζBFI [-]
r = 0.88
r = 0.88
r = 0.87
Observed ζHFD [days]
Simulated ζHFD [days]
r = 0.62
r = 0.66
r = 0.61
Observed ζQ5 [mm day−1]
Simulated ζQ5 [mm day−1]
Andelfingen
Appenzell
Frauenfeld
Halden
Herisau
Jonschwil
Mogelsberg
Mosnang
StGallen
Wängi

Figure 9: Simulated vs observed streamflow signatures. Single HRU model on the left, two HRUs model based on geology in the centre, two HRUs model based on land use on the right. Each dot represents a year and each colour a subcatchment. From up to bottom, mean daily streamflow (ζ_Q), baseflow index (ζ_{BFI}), mean half streamflow date (ζ_{HFD}), and 5th percentile of the streamflow (ζ_{Q5}). The red dashed line has a 45 ° slope and indicates where all points should align in case of perfect match. The Spearman's rank score (r) is also reported.

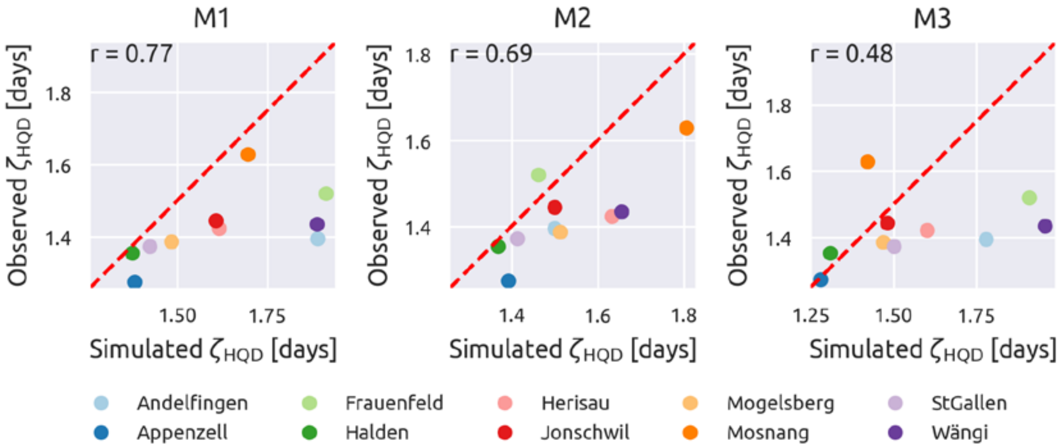


Figure 10: Ability of the hydrological models of representing the signature duration of low-flow events (ζ_{HQD}). Single HRU model on the left, two HRUs model based on geology in the centre, two HRUs model based on land use on the right.

Tables

Table 1: ~~**Landscape characteristics**~~**Identification of the** ~~**catchments. Only features that are not presented in**~~**gauging stations and description of the** ~~**plots are reported in this table**~~**river network.**

	Index	Code [a]	Upstream catchments
Andelfingen	1	2044	2 – 10
Appenzell	2	2112	–
Frauenfeld	3	2386	10
Halden	4	2181	2, 3, 5 – 10
Herisau	5	2305	–
Jonschwil	6	2303	7, 8
Mogelsberg	7	2374	–
Mosnang	8	2414	–
St. Gallen	9	2468	2
Wängi	10	2126	–

[a] Code of the gauging station, as defined by the Federal Office for the Environment FOEN

Table 2: Values of the streamflow signatures. The names of the subcatchments are abbreviated using the first three letters. The last column contains the coefficient of variation of each signature.

	Subcatchment										
	And	App	Fra	Hal	Her	Jon	Mog	Mos	StG	Wän	CV
ζ_{Q}	2.46	4.14	1.64	3.08	2.95	3.71	3.21	2.91	3.43	2.03	0.25
ζ_{RR}	0.63	0.80	0.49	0.70	0.71	0.80	0.70	0.72	0.71	0.56	0.14
ζ_{EL}	1.35	1.22	1.68	1.24	1.17	1.35	0.97	1.37	0.99	1.54	0.17
ζ_{FDC}	2.12	2.41	2.11	2.30	2.08	2.49	2.76	2.78	2.47	2.02	0.12
ζ_{BFI}	0.55	0.50	0.56	0.52	0.50	0.50	0.45	0.42	0.48	0.57	0.10
ζ_{HDF}	194.21	220.63	170.38	202.00	193.87	205.38	196.96	168.33	209.36	173.17	0.09
ζ_{Q5}	0.50	0.70	0.35	0.57	0.74	0.54	0.44	0.28	0.60	0.49	0.27
ζ_{Q95}	6.96	12.85	4.83	9.23	9.17	11.19	10.57	10.46	11.00	5.98	0.28
ζ_{HQF}	2.21	5.17	3.50	3.67	6.34	4.46	6.54	12.96	5.87	2.96	0.57
ζ_{HQD}	1.39	1.25	1.45	1.35	1.40	1.39	1.37	1.58	1.35	1.29	0.06
ζ_{LQF}	17.50	31.92	12.92	24.21	2.62	37.21	49.42	66.92	28.35	7.25	0.71
ζ_{LQD}	6.67	6.18	3.69	6.53	2.00	7.44	6.38	7.11	4.53	4.35	0.32

Table 3: Values of the climatic indices. The names of the subcatchments are abbreviated using the first three letters. The last column contains the coefficient of variation of each index.

	Subcatchment										
	And	App	Fra	Hal	Her	Jon	Mog	Mos	StG	Wän	CV
ψ_{P}	3.91	5.15	3.36	4.38	4.13	4.64	4.57	4.04	4.80	3.62	0.13
ψ_{PET}	1.60	1.37	1.70	1.55	1.61	1.54	1.57	1.69	1.49	1.71	0.07
ψ_{AI}	0.41	0.27	0.50	0.35	0.39	0.33	0.34	0.42	0.31	0.47	0.19
ψ_{FS}	0.04	0.21	0.04	0.05	0.09	0.15	0.13	0.09	0.13	0.05	0.57
ψ_{HPF}	15.21	14.38	17.67	14.58	15.82	14.54	14.58	16.13	14.31	17.50	0.08
ψ_{HPD}	1.20	1.17	1.17	1.18	1.22	1.20	1.19	1.22	1.17	1.19	0.01
ψ_{HDS}	Summer	Summer	Summer	Summer	Summer	Summer	Summer	Summer	Summer	Summer	0.00
ψ_{LPF}	201.67	195.79	216.83	198.54	205.04	197.21	198.92	205.75	197.69	213.17	0.04
ψ_{LPD}	3.57	3.50	3.83	3.50	3.63	3.51	3.51	3.66	3.51	3.76	0.03
ψ_{LPS}	Fall	Fall	Fall	Fall	Fall	Fall	Fall	Fall	Fall	Fall	0.00

Table 4: Values of the subcatchment characteristics. The names of the subcatchments are abbreviated using the first three letters. The last two columns contain the coefficient of variation and the maximum value of each signature.

	Subcatchment											
	And	App	Fra	Hal	Her	Jon	Mog	Mos	StG	Wän	CV	MAX
ξ_{A}	1701	74.46	213.34	1085	16.72	493.0	88.11	3.19	261.1	78.96	1.40	1701
ξ_{TE}	768	1250	591	908	831	1020	954	797	1039	650	0.22	1250
ξ_{TSm}	13.32	25.23	9.70	16.87	15.44	20.66	19.77	15.68	19.72	12.49	0.27	25.23
ξ_{TSs}	0.47	0.81	0.33	0.62	0.69	0.77	0.79	0.71	0.73	0.45	0.26	0.81
ξ_{TAs}	0.25	0.22	0.23	0.23	0.21	0.23	0.24	0.40	0.24	0.21	0.23	0.40
ξ_{TAn}	0.32	0.35	0.33	0.32	0.33	0.32	0.31	0.24	0.33	0.32	0.09	0.35
ξ_{TAew}	0.43	0.43	0.44	0.44	0.46	0.44	0.45	0.36	0.43	0.47	0.07	0.47
ξ_{SM}	1.30	0.56	1.48	1.10	1.32	0.93	1.17	1.00	1.03	1.35	0.23	1.48
ξ_{SD}	0.40	0.04	0.49	0.25	0.41	0.13	0.28	0.00	0.26	0.36	0.63	0.49
ξ_{LF}	0.26	0.25	0.28	0.27	0.21	0.31	0.34	0.18	0.27	0.30	0.17	0.34
ξ_{LC}	0.04	0.00	0.04	0.03	0.03	0.01	0.01	0.01	0.01	0.04	0.79	0.04
ξ_{LU}	0.08	0.03	0.10	0.06	0.15	0.04	0.03	0.03	0.05	0.10	0.63	0.15
ξ_{LP}	0.60	0.59	0.57	0.61	0.61	0.61	0.62	0.77	0.63	0.55	0.09	0.77
ξ_{GA}	0.06	0.01	0.09	0.03	0.00	0.02	0.02	0.00	0.01	0.11	1.05	0.11
ξ_{GC}	0.59	0.92	0.54	0.73	0.88	0.90	0.92	1.00	0.88	0.63	0.20	1.00
ξ_{GU}	0.35	0.07	0.36	0.23	0.12	0.07	0.06	0.00	0.10	0.26	0.79	0.36

~~**Table 2: Correlation coefficients (Pearson and Spearman's rank correlation) between streamflow signatures (columns) and meteorological and landscape characteristics (rows) of the catchments. Correlations that are statistically significant (p value < 0.05) are marked in bold in the table. Correlations that are interpreted to represent causality are underlined (this analysis has been done only for the Pearson correlation).**~~

~~**Table 3: Results of the linear regression with forward selection based on the results of the Pearson correlation. Each sub-table represents a hydrological signature and reports the coefficients of the regression and the evaluation metrics.**~~

Table A1: hydrological model parameters with range of variation used for the definition of the uniform prior distribution. The “component” column indicates the element (reservoir, lag or network) where the parameter belongs.

Parameter	Unit	Component	Range of variability
C_{e}	–	Unsaturated reservoir (UR)	0.1 – 3.0
$S_{\mathrm{max}}^{\mathrm{UR}}$	mm	Unsaturated reservoir (UR)	0.1 – 500.0
k_{WR}	d^{-1}	Snow reservoir (WR)	0.1 – 10.0
$t_{\mathrm{rise}}^{\mathrm{IL}}$	d	Network lag	0.5 – 10.0
$t_{\mathrm{rise}}^{\mathrm{OL}}$	d	Network lag	0.5 – 10.0
D	–	Structure	0.0 – 1.0
k_{FR}	d^{-1}	Fast reservoir (FR)	10^{-6} – 10.0
k_{SR}	d^{-1}	Slow reservoir (SR)	10^{-6} – 1.0
$t_{\mathrm{rise}}^{\mathrm{lag}}$	d	Structure lag	1.0 – 20.0

Table A2: Water–budget equations (see model schematic in Figure 6~~)~~).

Component	Equation
Snow reservoir (WR)	$\frac{\mathrm{d}S_{\mathrm{WR}}}{\mathrm{d}t} = P_{\mathrm{WR}} - Q_{\mathrm{WR}}$
Unsaturated reservoir (UR)	$\frac{\mathrm{d}S_{\mathrm{UR}}}{\mathrm{d}t} = P_{\mathrm{UR}} - Q_{\mathrm{UR}} - E_{\mathrm{UR}}$
Lag function	$Q_{\mathrm{UR}} = P_{\mathrm{SR}} + P_{\mathrm{lag}}$
Slow reservoir (SR)	$\frac{\mathrm{d}S_{\mathrm{SR}}}{\mathrm{d}t} = P_{\mathrm{SR}} - Q_{\mathrm{SR}}$
Fast reservoir (FR)	$\frac{\mathrm{d}S_{\mathrm{WR}}}{\mathrm{d}t} = P_{\mathrm{FR}} - Q_{\mathrm{FR}}$
Outflow	$Q = Q_{\mathrm{FR}} + Q_{\mathrm{SR}}$

Table A3: Constitutive functions of the model. Refer to Table A4 for the definition of the functions f. The calibrated parameters are marked in ~~bold~~red

Component	Equation
Snow reservoir (WR)[a]	$P_{\mathrm{WR}} = \begin{cases} P \text{ if } T \leq 0 \\ 0 \text{ if } T > 0 \end{cases}$
Snow reservoir (WR)[b]	$M_{\max}^{\mathrm{WR}} = \begin{cases} 0 \text{ if } T \leq 0 \\ k_{\mathrm{WR}} T \text{ if } T > 0 \end{cases}$
Snow reservoir (WR)	$Q_{\mathrm{WR}} = M_{\max}^{\mathrm{WR}} f_{\mathrm{e}}(S_{\mathrm{WR}}\|2)$
Unsaturated reservoir (UR)	$\overline{S_{\mathrm{UR}}} = \frac{S_{\mathrm{UR}}}{S_{\max}^{\mathrm{UR}}}$
Unsaturated reservoir (UR)	$E_{\mathrm{UR}} = C_{\mathrm{e}}(PET) f_{\mathrm{m}}(S_{\mathrm{UR}}\|0.01)$
Unsaturated reservoir (UR)	$Q_{\mathrm{UR}} = P_{\mathrm{UR}} f_{\mathrm{p}}(\overline{S_{\mathrm{UR}}}\|2)$
Slow reservoir (SR)	$P_{\mathrm{SR}} = D Q_{\mathrm{UR}}$
Slow reservoir (SR)	$Q_{\mathrm{SR}} = k_{\mathrm{SR}} S_{\mathrm{SR}}$
Lag function[c]	$P_{\mathrm{FR}} = (P_{\mathrm{L}} * h_{\mathrm{lag}})(t)$
Lag function	$h_{\mathrm{lag}} = \begin{cases} 2t/\left(t_{\mathrm{rise}}^{\mathrm{lag}}\right)^2 \text{ if } t \leq t_{\mathrm{rise}}^{\mathrm{lag}} \\ 0 \text{ if } t > t_{\mathrm{rise}}^{\mathrm{lag}} \end{cases}$
Fast reservoir (FR)	$Q_{\mathrm{FR}} = k_{\mathrm{FR}} S_{\mathrm{FR}}^3$
Lags in the network[c]	$Q_{\mathrm{out}} = (Q_{\mathrm{in}} * h_{\mathrm{lag}}^{\mathrm{net}})(t)$
Lags in the network	$h_{\mathrm{lag}}^{\mathrm{net}} = \begin{cases} 2t/\left(t_{\mathrm{rise}}^{\mathrm{OL/IL}}\right)^2 \text{ if } t \leq t_{\mathrm{rise}}^{\mathrm{OL/IL}} \\ \left(1/t_{\mathrm{rise}}^{\mathrm{OL/IL}}\right)\left(1-\left(\left(t-t_{\mathrm{rise}}^{\mathrm{OL/IL}}\right)/t_{\mathrm{rise}}^{\mathrm{OL/IL}}\right)\right) \text{ if } t_{\mathrm{rise}}^{\mathrm{OL/IL}} < t \leq 2t_{\mathrm{rise}}^{\mathrm{OL/IL}} \\ 0 \text{ if } t > 2t_{\mathrm{rise}}^{\mathrm{OL/IL}} \end{cases}$

[a] This equation is smoothed using logistic scheme, Eq. (8) in Kavetski and Kuczera (2007), with smoothing parameter $m_P = 1.5°C$

[b] This equation is smoothed using logistic scheme, Eq. (13) in Kavetski and Kuczera (2007)~~.~~, with smoothing parameter $m_M = 1.5°C$

[c] The operator $*$ denotes the convolution operator, smoothed according to Kavetski and Kuczera (2007)

Table A4: Constitutive functions

Function	Name
$f_{\mathrm{e}}(x\|\theta) = 1 - exp(-x/\theta)$	Tessier function. Note that $f_{\mathrm{e}}(x\|\theta) \rightarrow 1$ as $x \rightarrow \infty$
$f_{\mathrm{p}}(x\|\theta) = x^{\theta}$	Power function
$f_{\mathrm{m}}(x\|\theta) = \dfrac{x(1+\theta)}{x+\theta}$	Monod–type kinetics, adjusted so that $f_{\mathrm{m}}(1\|\theta) = 1$