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 structure of the paper, where we have separated the methods from the results, and the introduction, the discussion, and the conclusion sections where we tried to address the concerns of the reviewers. A summary of the changes is presented below, followed by the individual responses to the editor and to the reviewers.

Changes in text

SECTION	DESCRIPTION
Title	The title has been changed following the indications of Anonymous Referee #3.
1	The introduction has been changed adding more information about the objective of the study.
3	Section 3 now contains all the methodology, both of the correlation analysis and of the modeling study. Minor changes have been made to the single paragraphs.
4	Section 4 now contains all the results, both of the correlation analysis and of the modeling study. Minor changes have been made to the single paragraph.
5	The discussion has been enlarged, clarifying the choices made and pointing out their limitations.
6	Major changes to address the concerns of the reviewers.

Changes in Figures

The following figures were modified

FIGURE	DESCRIPTION
1	Different colors for the land use map.
6	Changed to address the concerns of Anonymous Referee #3.
7	Changed to address the concerns of Anonymous Referee #3.

Changes in Tables

Table 2 was added listing the signatures and the indices used in the study.

Reply to the editor Dr. Conrad Jackisch

Dear Marco Dal Molin and co-authors,

Thank you again for your contribution to our special issue and the work you invested into your manuscript's revision. After reading your manuscript again and considering the two independent reviewer reports, I follow their suggestion to open a second round of major revisions. Please be aware that both reviewers scored scientific quality and significance as "good" and presentation quality as "fair". Given your interesting material in your manuscript, I am sure this can benefit from considering the very thoughtful comments.

The two reviewers point out two lines of revisions. While Referee #4 has suggestions for fundamental clarifications of the study's aims and scopes. I read these not in a sense questioning your overall study but as valuable references a revision should orientate on. Hence I would expect that answering his questions on the fundamental level as key to structure the revisions on. Referee #3 addresses fundamental methodological aspects, which might on second look not be too far from the reflections of Referee #4. Considering the scope of our special issue as third pole, this lines up quite nicely in my view ("Linking landscape organisation and hydrological functioning: from hypotheses and observations to concepts, models and understanding").

If you see any trouble in addressing the comments during your revisions, please do not hesitate to contacting me for further clarification.

Thank you very much for your efforts and work you put into this manuscript.

All the best.

Conrad

We thank the Editor Dr. Conrad Jackisch for his thoughtful suggestions. We believe we have done another major review of the manuscript in order to address the comments of both reviewers. We noted that unfortunately the paper received a new set of reviewers, who came up with several new points, sometimes in contrast with the points raised in the first round of reviews. We did our best to address the comments of the current set of reviewers, without penalizing the changes already made to comply with the suggestions of the previous reviewers. Nonetheless, the paper underwent major changes, as can be noted in the differences' file, including a major restructuring, as suggested by Anonymous Referee #3.

We believe that our paper contributes to understanding the link between catchment properties and hydrological functioning, and therefore is well in line with the topic of the special issue. We are confident that the changes made, including the change in title and the new argumentations in the introduction, make this link even more visible.

Kind regards,

Marco Dal Molin (on behalf of the coauthors).

Reply to review by Anonymous Referee #3

We thank the reviewer for his/her careful read to the manuscript and insightful suggestions.
As it can be noticed in the differences' file, the paper has undergone a major restructuring, in the spirit of capturing most of the suggestions of the reviewers. However, as the reviewers in this round of reviews are different from the reviewers from the previous round of reviews, we had to be careful that the suggestions of the current reviewers are not in contrast with the modifications already made to comply to the suggestions of the previous reviewers. Cases where a conflict occurs are mentioned in our replies.
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.
All the references to specific pages and lines of the paper are based on the version without track changes. Since the numbering of the sections has changed in the reviewed paper, we will call "first revision of the paper" the version that you have reviewed and "new revision" the version that we are submitting together with this reply.

Main comments

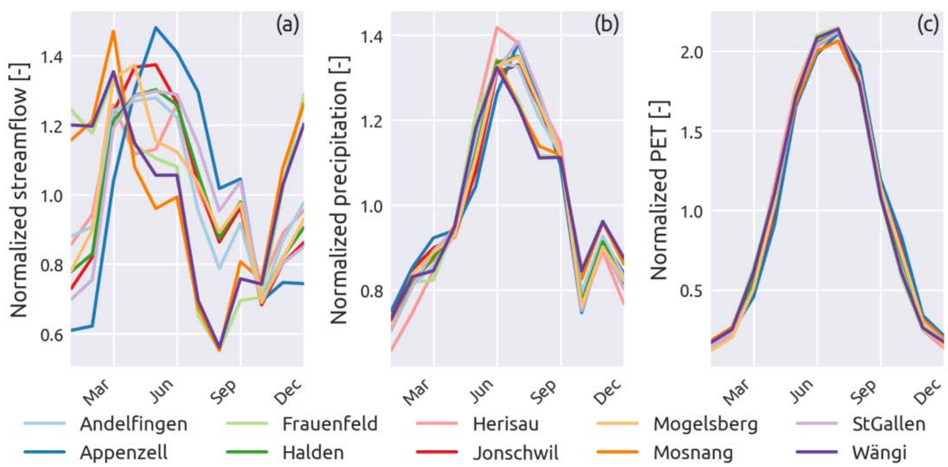
Dal Molin and colleagues submitted their revised manuscript entitled "Data analysis and model building for understanding catchment processes: the case study of the Thur catchment" to Hydrology and Earth System Sciences (HESS) Special Issue: Linking landscape organisation and hydrological functioning: from hypotheses and observations to concepts, models and understanding. The manuscript was substantially improved after major modifications following first iteration with reviewers and editors and I enjoyed reading it. However, I failed to identify a major scientific contribution in terms of processes understanding supported by hydrological interpretations, which makes me feel that authors are targeting the proposition of the regional modelling framework rather than the potential hydrological insights of the modelling exercise. In that sense, there are some issues that need to be addressed in order to achieve a replicable regional modelling framework, which are discussed in details below.

1. *Section 3 still worries me a bit. First, the climate indices presented in Addor et al (2017) were selected to be representative in large-scale studies (e.g., CAMELS) where a large climatic gradient is the main control of catchment's streamflow spatial variability. This is not necessarily a valid assumption for regional studies where climate variability (across space) is much smaller and variables are highly correlated (as per Fig 3).*
   The reviewer is right saying that in a limited region, such as the one presented in this study, the climatic conditions may not vary as strongly as in studies targeting large climatic gradients (e.g. CAMELS); however assuming no climatic variability in the Thur catchment would be unjustified: for example, the mean annual precipitation varies significantly between the subcatchments (e.g. 5.15 mm/d in Appenzell vs. 3.36 mm/d in Frauenfeld); moreover the variability of the amount of precipitation falling as snow is large (e.g. 21% in Appenzell vs. 4% in Frauenfeld) which induces significant differences in streamflow seasonality, as we have shown in figure 5 presented in the

first submission of the paper, then removed to comply to the reviewers suggestions, and reported below.
The figure shows that, while precipitation and potential evapotranspiration follow the same annual pattern in all the catchments, the streamflow follows two different patterns, dividing the catchments in two different groups:

- Snow affected catchments (e.g. Appenzell) with high streamflow during the late spring and summer;
- Catchments with less snow (e.g. Frauenfeld) with highest streamflow between October and March.

Since this behavior was also captured by the "mean half streamflow date" signature, the figure was omitted by the second submission of the paper, as suggested by Anonymous Referee #1.



Second, the correlative nature of analysis underpinning decisions in the process-based hydrological modelling is a weakness. A more comprehensive metrics assessment would benefit the manuscript (see second point below).
We will address this point below
Third, the final "expert judgment" adds subjectiveness and undermines replicability of modelling framework. The other way around would be more intuitive – i.e., run the "expert judgment" prior in order to select relevant metrics and establish a process-based conceptual (see third point below) model taking accounting relevant specificities of study area (see point 4 below) and then the metrics assessment part.
In principle there are many climate and landscape characteristics that influence catchment response. The question is which one are the most relevant for the application, in particular at the spatial scale of the study and for the variables that one wants to predict. The reason for running the expert judgment after the correlation analysis is to be able to derive some of these key model decisions from it and not to decide them a-priori which would be difficult, if not impossible.

2. *Highly correlative nature of section 3: in the metrics assessment part (section 3), essentially, criteria for metric selection should go beyond correlation and represent similarity, dissimilarity, complementarity and/or importance of metrics and uncertainty. Bray-Curtis ordination or PCA could be helpful to understand data structure and complementarity and random forests could be used to calculate importance of metrics (see Kennard at al (2010) River Res Applic and Trancoso et al (2016) JoH for analytical examples). That would strength the analytical component of section 3.*

   The possibility to use more advanced methods for metrics selection has been considered in the process of our study; the reason why we eventually selected a simple method is that the sample size of this study is relatively small. We are in fact limited to only 10 catchments. Studies that use complex regression techniques like random forests use a much larger sample of catchments; for example, the work proposed by Trancoso et al. (2016) deals with 355 catchments. Using such techniques risks to result in models that overfit the data, especially considering the fact that we would need to split the catchments in a calibration and a validation group.

   In the previous revision of the paper, we took the suggestion of the first set of reviewers and used Spearman correlation instead of Pearson, in order to account also for nonlinear correlation.

   We have been more specific about the reasons behind this choice in the "limitations" part of the discussion.

   In the attempt to comply with the suggestion of the reviewer, we tested LASSO regression using catchment characteristics and climatic indices (indicated as x_i) to predict every single hydrological signature (indicated as y).

$$y = \sum \lambda_i x_i + \alpha \sum |\lambda_i|$$

   The idea behind this method is that it should perform (thanks to the normalization term $\alpha \sum |\lambda_i|$) feature selection, setting to zero all the λ_i that are associated with indices that are not important in calculating the output y.

   While being, in theory, a technique suitable also for a small unbalanced (unbalanced in the sense that there are more indices than catchments) data set, in the case of the Thur catchment the method performed poorly mainly because of the high correlation among the indices. In particular, the method prefers the most correlated variables, but we may be interested in a slightly less correlated variable if this reflects a more plausible cause effect relationship.

   One possible solution would be applying a pre-selection of the indices based on expert judgment (as done in section 3.2.1 of the first revision of the paper) before doing LASSO regression but this would fall back to the fallacies in the methodology criticized by the reviewer. If a pre-selection is done, this method would produce results comparable with the one given by the correlation analysis (done in section 3.2.2 of the first revision of the paper), with the disadvantage that the information that we get from the LASSO regression would be only a list of selected indices, while the correlations express also the strength of the relationship between indices and signatures.

   For these reasons, we have preferred to keep using the (non-linear) correlation analysis, aided by expert judgment, to select the meaningful catchment characteristics and climatic indices that influence streamflow signatures: it is true that this approach may be subjective but it guaranties meaningful insights for building the hydrological model that are not affected by spurious correlations.

Uncertainty in the streamflow signatures and in the climatic indices was not considered because the time window used for their calculation (24 years) is long enough to assume limited bias and high precision (as shown by Kennard et al. (2010)).

3. *A conceptual model would be helpful on section 3 to understand how selected metrics represent catchment processes.*
   Point taken. Figure 6 has been modified showing how the catchment has been modeled.
4. *Metrics from continental scale studies are a good starting point but should not be the final call – there are also other relevant metrics that could be tested such as phase-offset between the seasonal cycle in precipitation and potential evaporation – see Donohue et al (2010) JoH for details.*
   The number of signatures and indices proposed in literature to represent streamflow and climate is enormous (e.g. 120 metrics considered by Kennard et al. (2010)); therefore we had to limit our selection and we decided to use the one proposed by Addor et al. (2017) since we think they cover a wide range of characteristics of the time series that they synthetize.
   In the first submission of the paper, the climatic indices and the streamflow were selected to represent particular features of the time series (e.g. the flashiness index was used to measure the variability of the hydrograph). This choice was criticized in the review of Dr. Lieke Melsen that suggested to refer to other studies for the selection of signatures/indices. Therefore, the original version of the paper was modified to account for her suggestion.
   We have also tried to calculate the phase-offset between the seasonal cycle in precipitation and potential evapotranspiration as suggested by this reviewer but, since precipitation and PET have the same seasonal cycle in all the catchments (as shown in figure 5 of the first submission of the paper, reported above), the phase-offset would be the same for all the catchments and therefore it would be excluded by the correlation analysis since it does not show variability (first bullet point in section 3.1.2 of the new revision). For this reason we would not include it in the paper.
5. *Manuscript structure is a bit unusual and not easy to navigate. It looks more like a thesis/report than a paper. Would be better to group all the methods and results together instead of presenting them separately on sections 3 and 4.*
   We thank the reviewer for this suggestion. Now the paper has been restructured according to the standard practice of presenting methods followed by results. We hope that this major restructuring has improved the readability of the paper.

Other comments

Manuscript title is focused on methods rather than contribution. Currently title is a bit vague and not attractive as most hydrological modelling papers do data analysis and model building to understand catchment processes. Therefore, there is nothing new in the title and many potential readers might skip it if the title is kept the same (It is likely I would be one of them). I strongly recommend changing it focusing on the main contribution.

Point taken. We have changed the title into: “Understanding dominant controls on streamflow spatial variability to set-up a semi-distributed hydrological model: the case study of the Thur catchment”, which focuses more clearly on the paper objectives.

L1-2: “The development of semidistributed hydrological models that reflect the dominant processes controlling streamflow spatial variability is a challenging task” – Irrelevant, every science has challenges, otherwise would not be science.

We have removed the sentence from the new version of the paper, and added two paragraphs in the introduction to better clarify the scope of our work.

The term “semidistributed” sometimes appears as semi-distributed. Better standardise.

Thank you for pointing that out; we have standardized this term in the paper, using semi-distributed everythere.

Figure 1c: Forest and pasture are not easy to distinguish. Suggest use different colour for pasture.

We have changed the colors used in the figure. Now the figure should be more readable.

Figure 7 uses line plots to show variability of model performance across study catchments. Choice of plot type is a bit misleading as line plots are usually used to show continuity across the x-axis, such as time-series plots. Suggest use only dots / jitters instead to avoid misinterpretation.

We have removed the lines from figure 7

Most figures refer to metrics acronyms and a lot of back and forth is needed to find their definition and keep on track with reading. If authors do not want to redefine acronyms on figure captions, suggest present all the metrics in a table and cite table on figure captions.

We have added table 2 that contains all the symbols used in the paper to represent signatures and indices

P28 L2 – “varied considerably between catchments” – I think it should be among instead of between as several catchments are assessed.

We have changed the sentence accordingly.

P28 L7-8: “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” – that’s well known and trivial. I would expect more elaborated hydrological insights. Main issue is that section 3 is not robust enough to offer in-depth interpretations.

Although it is known that precipitation has a strong control on average streamflow, this is not granted in some cases where, for example, regional groundwater flow alters the water balance of the catchments. For this reason we believe that this point is still important for describing the hydrological processes happening in the catchment.

We have clarified these aspects in Section 4.1.3 point 1, Section 4.2.1 first paragraph, and Conclusions, point 4 of the new revision.

We are thankful for the reviews of Anonymous Referee #3 and we are looking forward to his/her assessment of our revised paper.

References

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.

Donohue, R. J., Roderick, M. L., & McVicar, T. R. (2010). Can dynamic vegetation information improve the accuracy of Budyko's hydrological model?. Journal of Hydrology, 390(1-2), 23-34.

Kennard, M. J., Mackay, S. J., Pusey, B. J., Olden, J. D., & Marsh, N. (2010). Quantifying uncertainty in estimation of hydrologic metrics for ecohydrological studies. River Research and Applications, 26(2), 137-156.

Trancoso, R., Larsen, J. R., McAlpine, C., McVicar, T. R., & Phinn, S. (2016). Linking the Budyko framework and the Dunne diagram. Journal of Hydrology, 535, 581-597.

Reply to review by Dr. Shervan Gharari

We thank Dr. Shervan Gharari for his careful read to the manuscript and insightful suggestions. As it can be noticed in the differences' file, the paper has undergone a major restructuring, in the spirit of capturing most of the suggestions of the reviewers. However, as the reviewers in this round of reviews are different from the reviewers from the previous round of reviewers, we had to be careful that the suggestions of the current reviewers are not in contrast with the modifications already made to comply to the suggestions of the previous reviewers. Cases where a conflict occurs are mentioned in our replies.
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.
All the references to specific pages and lines of the paper are based on the version without track changes. Since the numbering of the sections has changed in the reviewed paper, we will call "first revision of the paper" the version that you have reviewed and "new revision" the version that we are submitting together with this reply.

Review of "Data analysis and model building for understanding catchment processes: the case study of the Thur catchment"
The paper tries to rationally build/infer an appropriate model suture based on data for Thur catchment. The authors have tried their best to answer to the reviewers' comments. Reading the manuscripts, reviewers' comments on the work and authors' replies, I have a feeling that most of the reviewers' concerns including myself is coming from the fact that the manuscript lack some fundamental direction which in turn might be the result of lack of proper research question. I personally don't have any issue with the choice of signatures and correlation analysis. At the end of the day this is an engineering decision that any modeller will make and there is little to back them even if correlation exists (present or absence of causation).
The first question the authors should answer is the real purpose of this study.
We have modified abstract and introductions to clarify the purpose of the study. In particular, the first paragraph of the introduction presents the general purposes of conceptual semi-distributed hydrological models in hydrology and some unresolved questions, which now better substantiate objectives of the studies, indicated in lines 37 of page 2 to 2 of page 3. The title has also been changed to more clearly reflect the study objectives, as suggested also by Anonymous Referee #3.

The spatial variability can have a wide range of interpretation. For example spatial variability to streamflow, or spatial variability to account for slope and aspect and etc. The authors should clearly make this case what spatial variability they are talking about (variability is case dependent).
To avoid misunderstandings, we have clarified that we are interested in explaining the hydrograph spatial variability. This is now more clearly apparent from the title ("Understanding first order controls on streamflow spatial variability…"), the abstract (e.g. "In order to appraise the dominant controls on streamflow spatial variability, and build a model that reflects them…"), and objectives (e.g. "The objective of this study is to develop a semi-distributed hydrological model with the appropriate level of functional complexity to reproduce streamflow spatial variability in the Thur catchment."). If the

catchment response is spatially variable there must be some spatially variable controls, and therefore we also analyze the spatial variability of meteorological inputs and catchments characteristics.

In this study, it is all about the streamflow as the models are calibrated against the streamflow. Streamflow is often easy to predict (calibrate). So the author should show the clear gain by moving toward spatially distributed input and spatial data such as slope and aspect, vegetation, geology, etc. In its current format the manuscript is lacking this direction. In the beginning, the manuscript promises to account for various processes but it is kind of missed in the manuscript or boiled down to very basic or common knowledge interpretation of the processes for streamflow simulation.

We hope our restructuring of the manuscript, where the methods are all presented in the same place, makes the reasoning of the paper clearer. In particular, we have revised sections 3.1.4 and 4.1.3 of the new revision to clarify these aspects.

In summary, the starting point for the modeling study is a semi-distributed model with uniform characteristics (single HRU) and distributed (per subcatchment) climatic inputs (M0) as the effect of distributing precipitation would be obvious from the signatures analysis; we then show the gain in moving towards accounting for the presence of spatially distributed snow, geology, and land use (vegetation).

As the reviewer notes, there are multiple characteristics that could be included in the model experiments. In order to reduce the number of model comparisons, we made use of the results of the signatures analysis. For example, signatures analysis showed that vegetation was not a major influencing factor and, in our model experiments, we confirmed that including vegetation does not improve model performance. Similarly, it could be expected that accounting for e.g. aspect, which was not a major influence factor according to the signatures analysis, would not improve model performance.

Sth else that I don't understand is the choice of model, for example from M0 to M1, if temperature is always above the threshold there will be no phase change for the precipitation. Then why even bother having model M0? The choice of the model is decided by the data itself (for example a land surface models have always snow component but if simulated for warm region they never simulate any snow).

It is clear that when there is snow (as in this case) the model needs to have a snow component. It is less obvious (at least just by looking at hydrographs) how much of the differences in seasonality of the streamflow response between catchments are due to snow. Due to the large lag time between snowfall and hydrograph response it would be difficult to quantify this aspect without model experiments and the main result of the comparison between M0 and M1 is that the attribution of difference in seasonality (represented by the mean half streamflow date) is due to the spatial variability of snow processes.

Moreover, the fact that the precipitation is first order control is also a bit obvious. If a multiplier is used to scale precipitation up and down it will be the most sensitive parameter of the model which in turn shows that the simulation is heavily affected by precipitation (or the driving force). Following that, I don't see much translation of the observed processes into the model and I don't see the added value of the added heterogeneity in the spatial models simulation etc. This can be further improved by the authors.

It is clear that precipitation is a first order control on streamflow. Less clear, at least before carrying out any analysis, is if the spatial variability in streamflow average is only due to precipitation: several authors, for example, pointed out the role of regional groundwater flow and incorporated this possibility in the models; GR4J, for example, has a parameter that quantifies catchment gains or losses. This shows that a-priori there are several processes that can affect the water balance; our analysis is intended to understand which modeling decisions are relevant in this case study. We have clarified these aspects in Section 4.3 point 1, Section 4.2.1 first paragraph, and Conclusions, point 4 of the new revision.

Following this point, the choice of the models and the modeling looks a bit sloppy; in the sense that the continuum of model, spatial data is not very well visible. I think this can be further improved by the authors in the revised manuscript (maybe adding more model or stepwise introduction of spatial variability).

Because spatially distributed models are time consuming to develop, even within a multi-model framework, and expensive to run, we focused the comparisons on a few interesting cases which were decided following the signatures analysis. We have clarified the models line-up in Section 4.2.1 of the new revision, where we have specified the expectations that the various models are supposed to meet.

I also suggest the author to have a more structural in to the paper by organizing the signatures that they use for model evaluation. These signatures can be grouped into four main categories (1) the signatures that are coming from the spatial heterogeneity of the topography, geology, soil, land cover etc. (2) the signature that are coming from the response that the model is built to replicate such as flow duration curve, flashiness, etc, (3) the signatures that are coming from the forcing to the system including the precipitation etc (4) hybrid such as runoff ratio. Each of these signatures have their own effect on the modeling result as some are used for calibration and some are not. I would suggest the author to segregate them more carefully in the test and analyses.

We tried to be explicit about the different nature of these metrics: we called metrics derived from the landscape properties “landscape indices” and indicated them with the letter ξ; the metrics derived from the climate were named “climatic indices” and indicated them with the letter ψ; the metrics derived from the streamflow, finally, were named “streamflow signatures” and indicated them with the letter ζ. Only the runoff ratio and the streamflow elasticity are “hybrid signatures” and we have decided to put them in the category of the “streamflow signatures” as done by Addor et al. (2017).

Is it surprising M0 can do well for the annual average? In my experience, a single reservoir with evapo-transpiration function can get the annual mean streamflow perfectly well, while it cannot get the correlation and variability.

The reviewer is right pointing out that a simple model calibrated on an individual gauge can get the average streamflow correctly; however this is a distributed model which is simultaneously calibrated on multiple gauges without catchment specific correction factors for precipitation, evaporation or streamflow. In addition, the model is evaluated in space-time validation, meaning that the model has not been calibrated in the specific gauge where it is evaluated. The ability of this model to simultaneously capture the annual averages at multiple gauges is, therefore, not a-priori obvious.

I am interested to know how the authors dealt with the nested gauges. The information/correlation in nested gauges can be replicated. Howe the correlation plays in for these nested basin. Any comment on that.

The model deals with nested catchments by routing the water from upstream catchments to the downstream outlets through transfer functions. By validating the model in space and time we are not reusing the same data. Clearly, this is a spatial validation in a nested setup, which is presumably easier to fulfill than a spatial validation in entirely different basins. This limitation has been added in Section 5 of the new revision, second last paragraph.

Why did the authors have use NS and likelihood at the same time? What would it add…?

We used the likelihood because it was the objective function for model calibration. Since the model was calibrated simultaneously in multiple stations, the likelihood is an aggregated metric. The NS was, on the other hand, calculated for each catchment individually.

I think both in the modeling set up and also discussion a significant elements regarding the scale is missing. For example, have refereed to some work, Kuentz et al., 2017, that did a large sample hydrology. Is the manuscript really is about large sample hydrology and if the study area is following large sample hydrology or is it about how the Thur catchment is functioning and how it is modelled.

We have clarified that the paper is not about large sample hydrology but about distributed modeling (e.g. see first paragraph of the introduction). In order to formulate model decision we have used elements of catchment classification studies.

I would also suggest the author to look into the signature before and after bias correction or accounting for orographic effects. As mentioned earlier changing in forcing can drastically change the model output therefore it should also be noted how different the forcing becomes when is downscaled. Maybe I missed but how did the author include slope and aspect in their model?

Climatic inputs are influenced by orographic effects; as specified in Section 2 of the paper, the elevation has been considered in the interpolation of the data from the meteorological stations. We did not do sensitivity analyses on the input variables as it was outside of the scope of this paper.

I believe the manuscript can be an interesting contribution but in its current format it is far from being in perfect shape. The story needs to follow smoothly and the merit of this work should be better presented.
Shervan Gharari

We are thankful for the reviews of Dr. Shervan Gharari and we are looking forward to his assessment of our revised paper.

References

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.

~~Data analysis and model building for understanding catchment processes~~Understanding dominant controls on streamflow spatial variability to set-up a semi-distributed hydrological model: 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~~ This study documents the development of ~~semidistributed~~a semi-distributed hydrological ~~models that reflect~~model aimed at reflecting the dominant ~~processes controlling~~controls on observed streamflow spatial variability ~~is a challenging task. This study illustrates this~~. The process is presented through the case study of the Thur catchment (Switzerland, 1702 km^2), an alpine and pre–alpine catchment ~~with large spatial variability in terms of climate, landscape, and~~ where streamflow (measured at 10 subcatchments~~).~~) has different spatial characteristics in terms of amounts, seasonal patterns, and dominance of baseflow. In order to appraise the dominant ~~processes that control catchment response~~controls on streamflow spatial variability, and build a model that reflects them, ~~the model development follows~~we follow a two–stages approach. In a first stage, we ~~use correlation analysis to~~ identify the main ~~influencing factors on~~ climatic or landscape properties that control the spatial variability of streamflow signatures. ~~Results of this analysis show that precipitation averages control signatures of water balance, snow processes control signatures of seasonality, while landscape characteristics (especially geology) control signatures characterizing the importance of baseflow.~~This stage is based on correlation analysis, complemented by expert judgment to identify the most plausible cause-effect relationships. In a second stage, the results of the previous analysis are used to develop a set of model experiments aimed at determining an appropriate model representation of the Thur catchment. These experiments confirm that only a hydrological model that accounts for the heterogeneity of precipitation, snow related processes, and landscape features such as geology, produces hydrographs that have signatures similar to the observed ones. This model provides consistent results in space–time validation, which is promising for predictions in ungauged 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 numerous regions around the globe.

1 ~~1~~ Introduction

~~Due to the spatial variability of landscape (e.g. topography, land use, etc.) and climate, 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 between places,~~Semi-distributed rainfall-runoff models are widely applied in operation for applications such as flood forecasting (e.g., Ajami et al., 2004) or developing sustainable irrigation practices (e.g., McInerney et al., 2018). The main purpose of these models is to simulate streamflow at a limited number of fixed points along river channels (e.g., Boyle et al., 2001), and for this reason they are characterized by a coarser spatial resolution than fully distributed models, which allows a very detailed representation of the spatial variability of catchment processes. Compared to fully distributed models, they are characterized by lower data and computational requirements, which represents an valuable practical advantage in their operational use.

Similarly to the case of lumped models, the parameters of semi-distributed models are estimated via calibration. Therefore, it is important that the structure of these models is commensurate to the available data, including length, time scale, and spatial distribution (Wooldridge et al., 2001). However, semi-distributed models used for similar applications differ significantly in terms of their process representation as well as number of calibration parameters. For example, Gao et al. (2014) assumes topography as a dominant control on hydrological processes, whereas the SWAT model (Arnold et al., 1998) emphasizes the role of soil. These differences raise the question of how to select an appropriate model structure for the data at hand, which requires understanding the association between model parameters and the climatological and geomorphological characteristics of the catchment.

Understanding the relationship between climate, landscape and catchment response 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)~~(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)~~(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 are designed to regionalize selected characteristics of the hydrograph (streamflow signatures), or through hydrological models that 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 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 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~~forcings and landscape characteristics, ~~semidistributed~~semi-distributed hydrological models have the ability to 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 many parameters. For example, Gurtz et al. (1999) considered several landscape characteristics (elevation, land use, etc.) in their application of a ~~semidistributed~~semi-distributed 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, 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. ~~Antonetti et al. (2016)~~Antonetti et al. (2016) used a map of dominant runoff processes following ~~Scherrer and Naef (2003)~~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).~~(e.g., Andréassian et al., 2009). A common approach for calibration of ~~semidistributed~~semi-distributed 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. 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 semidistributed hydrological models by first using 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~~The objective of this study is to develop a semi-distributed hydrological model with the appropriate level of functional complexity to reproduce streamflow spatial variability in the Thur catchment. For this purpose, we use a two stages approach, the first one dedicated to an in-depth analysis of the available data, and the second one focused on hydrological modeling.

Our specific objectives are to: (1) explore the spatial variability present in the Swiss Thur catchment regarding landscape characteristics, meteorological forcing and streamflow signatures; (2) identify the main ~~drivers~~climate and landscape controls that explain the variability of the hydrological response; (3) based on this analysis, build a set of model experiments aimed to test the relative importance of dominant processes and their effect on the hydrograph; (4) 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 availability; Section 3 ~~and Sect. 4 are both divided in methods and results and present, respectively,~~illustrates the ~~correlation analysis and~~methodology; Section 4 shows the ~~modeling part of this paper~~results; Section 5 ~~puts~~analyzes the results ~~of this work~~ and puts them in perspective, ~~comparing them with~~showing what other similar studies have found; Section 6, finally, summarizes the main conclusions.

2 ~~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. The Thur river is very dynamic, with streamflow values that can change by two orders of magnitude within a few hours ~~(Schirmer et al., 2014)~~(Schirmer et al., 2014).

The Thur catchment has been subject of several studies in the past; Gurtz et al. (1999) performed the first modelling study on the entire catchment using a 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~~who investigated parameters estimation in data limited scenarios and their transferability across spatial and temporal scales, and ~~Brunner et al. (2019)~~Brunner et al. (2019) 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, which presents a mountainous landscape.

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 7 % and 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 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)~~(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 (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 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 resolution in the subsequent analyses, aggregating the available hourly data. This choice was influenced on the 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)~~(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)~~(e.g., Kirchner et al., 2004).

3 Methods

~~3~~ The methodology follows a two stages approach. The first stage aims at determining the climatic and landscape controls on streamflow signatures. The second stage uses this understanding to configure the structure of a semi-distributed model, whose functional suitability is tested through a set of model experiments. Section 3.1 describes the first stage of the analysis, that is, the identification of influencing factors on the spatial variability of streamflow signatures. Section 3.2 describes the general structure of the semi-distributed model, and the model evaluation approach. The design of the model experiments, which is dependent on the outcomes of the first stage of analyses, is presented directly in the results (Sect. 4.2.1).

~~2.1~~3.1 Identification of influencing factors on the spatial variability of streamflow signatures

~~2.2 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, potential evaporation, and temperature time series. Landscape characteristics are represented by maps of topography, land use, geology, and soil.

Climatic conditions, landscape characteristics and streamflow are represented through a set of statistics~~.~~ (listed in Table 2). In the following, statistics calculated based on streamflow data will be called streamflow "signatures", as it is often done in catchment classification literature (e.g., Sivapalan, 2006). We will refer to climatic and landscape "indices" for statistics calculated based 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~~the 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; Sect. 3.1.4 describes how the signature analysis is used to set-up the model experiments.

~~2.2.1~~3.1.1 ~~3.1.1~~ Catchment indices for representing streamflow, climate, and landscape

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

~~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~~ $\cancel{\zeta_{\mathrm{Q}} = \overline{\boldsymbol{q}}}$~~, where~~ $\cancel{\boldsymbol{q}}$ ~~is the streamflow time series and the overbar represents the average over the observation period;~~
- ~~runoff ratio~~ $\cancel{\zeta_{\mathrm{RR}} = \frac{\overline{\boldsymbol{q}}}{\overline{\boldsymbol{p}}}}$~~, where~~ $\cancel{\boldsymbol{p}}$ ~~is the precipitation time series;~~
- ~~streamflow elasticity (~~$\cancel{\zeta_{\mathrm{EL}}}$~~) defined as~~

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

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: while being originally introduced for a study about large sample hydrology, we believe that the indices proposed are also able to capture several different aspects of the time series and are therefore suitable also for this regional study. The streamflow signatures here considered are described hereafter, followed by an explanation of their rationale:

- average daily streamflow $\zeta_{\mathrm{Q}} = \overline{\boldsymbol{q}}$, where $\boldsymbol{q}$ is the streamflow time series and the overbar represents the average over the observation period;
- runoff ratio $\zeta_{\mathrm{RR}} = \frac{\overline{\boldsymbol{q}}}{\overline{\boldsymbol{p}}}$, where $\boldsymbol{p}$ is the precipitation time series;
- streamflow elasticity (ζ_{EL}) defined as

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

where $\Delta\overline{\boldsymbol{q}}$ and $\Delta\overline{\boldsymbol{p}}$ represent the streamflow and precipitation ~~jumps~~difference 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 $\zeta_{\mathrm{BFI}} = \frac{\overline{q^{(\mathbf{b})}}}{\overline{q}}$, where $\boldsymbol{q}^{(\mathbf{b})}$ represents the baseflow and was calculated using a low–pass filter as illustrated in ~~Ladson et al. (2013)~~Ladson et al. (2013) with the equation

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

$$q_{\mathrm{t}}^{(\mathrm{b})} = q_{\mathrm{t}} - q_{\mathrm{t}}^{(\mathrm{f})} \quad (3)$$

with $q_{\mathrm{t}}^{(\mathrm{f})}$ 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 date (ζ_{HFD}) ~~(Court, 1962)~~(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;
- 5th and 95th percentiles of the streamflow ($\zeta_{\mathrm{Q5}}$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)~~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):~~

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

- average daily precipitation $\psi_{\mathrm{P}} = \overline{\boldsymbol{p}}$;
- average daily PET $\psi_{\mathrm{PET}} = \overline{\boldsymbol{e}_{\mathbf{pot}}}$, where $\boldsymbol{e}_{\mathbf{pot}}$ is the potential evapotranspiration time series;
- aridity index $\psi_{\mathrm{AI}} = \frac{\overline{\boldsymbol{e}_{\mathbf{pot}}}}{\overline{\boldsymbol{p}}}$;
- fraction of snow (ψ_{FS}), defined as the volumetric fraction of precipitation falling as snow (i.e. on days colder than 0 °C);
- frequency (ψ_{HPF}) and mean duration (ψ_{HPD}) of high precipitation events: they are defined as days when the precipitation is bigger than five times the 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 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)~~Addor et al. (2017) was not considered in this study as it relied on fitting a sinusoidal function to the 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~~subcatchment, 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 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) was defined. It is important to notice that all the areas calculated in this analysis were normalized by the respective subcatchment area (ξ_{A}) in order to get comparable values between subcatchments of different size.

Topography was represented with the following indices, calculated based on the digital elevation model (DEM):

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

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

- fraction of the subcatchment with crops land use (ξ_{LC});
- fraction of the subcatchment with pasture land use (ξ_{LP});
- fraction of the subcatchment with forest land use (ξ_{LF});
- fraction of the subcatchment with urbanized land use (ξ_{LU}).

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) (ξ_{SD});
- average soil depth (ξ_{SM}).

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 (ξ_{GA});
- fraction of the subcatchment with consolidated geology (ξ_{GC});
- fraction of the subcatchment with unconsolidated geology (ξ_{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 their number, in order to facilitate subsequent analyses.

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~~semi-distributed 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).

~~2.2.2~~3.1.2 ~~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~~streamflow 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~~assessed 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~~This 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 considered. Here, landscape indices accounting for less than 5 % of the subcatchment area were discarded.
- Within each set of streamflow signatures, climatic indices, and catchment indices we retained only relatively independent metrics~~.~~, if these are believed to represent the same underlying features of the time series. 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 absolute value of the 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.

~~2.2.3~~3.1.3 ~~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~~;~~.

The outcome of this process will be used to inform the semi-distributed model setup. The expert judgment is a critical step in the elicitation of causality from correlation (e.g., Antonetti and Zappa, 2018), and it is clearly subjective, being dependent on personal experience and subject matter knowledge. Although personal and subjective, expert decisions are based on an attempt to interpret data rather than be a-priori defined, which is typically the case in the application of semi-distributed hydrological models.

3.1.4 Semi-distributed model setup and model experiments

We assumed a generic structure for a semi-distributed hydrological model, described in Section 3.2.1, where some model structure characteristics are defined a priori, and others are to be defined. In order to motivate the open decisions, we proceeded as follows:

- 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: all these choices were, in this study, motivated by the results of the correlation analysis, i.e. only catchment characteristics that were found capable of explaining the hydrological response were used.~~

~~31~~ ~~3.2~~ interpret~~Results and interpretation~~

- ~~This section illustrates~~ the ~~results of the correlation analysis aimed to identify~~dominant processes influencing ~~factors that control the~~signature spatial variability~~ of streamflow signatures; Section 3.2.1 presents the 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.~~;

~~3.1.11.1.1~~ ~~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 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, a subset of~~ ζ ~~can be defined as follows:~~

- 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}~~), 5th 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.~~

~~3.1.21.1.1 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 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 sufficient to reach a p-value lower than 0.05.~~

~~These results are the premise for designing meaningful model experiments.~~

~~3.1.31.1.1~~ ~~3.3~~ designed model experiments aimed to confirm the hypothesized climatic and landscape controls on streamflow~~Hypotheses for model building~~

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

1. ~~The precipitation is the first driver of the differences in the water balance of the subcatchments. The effect of topographic variability manifests itself primarily as an influence on precipitation (amount and type). Accounting for variability of precipitation therefore implicitly reflects such effect of topography on the hydrograph, since the inputs were interpolated taking into account the effect of the elevation (Sect. 2).~~
2. ~~2.~~1. ~~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. ~~4.~~1. ~~The other catchment characteristics (e.g. soil, vegetation) show little or no correlations with the streamflow signatures and therefore they should not be considered if the idea is to keep the model as simple as possible.~~

~~These hypotheses will be tested through specific model comparisons, described in Sect. 4.1.5.~~

- 4

The overall objective of the model experiments is to prove that only models that incorporate the correct dependencies are able to correctly predict regional streamflow variability. In order to test this assumption, the model experiments will include cases where the assumed dependencies are not incorporated. Omitting an assumed dependency leads to structurally simpler model, which may raise the doubt that potential differences in model performance might be due to differences in model complexity. For this reason, the model experiments will include cases where alternative dependencies are incorporated, which do not reduce model complexity. In order to keep the study and presentation tractable, the model experiments will be limited to a few cases, illustrated in Sect. 4.2.1 which we judge relevant for this specific application.

3.2 General structure of the semi-distributed hydrological model and model evaluation approach

~~3.21.1 Modelling~~

~~41 4.1 Methods~~

This section describes the approach for building and testing a semi-distributed hydrological model designed to represent the observed streamflow and particularly the observed spatial variability of streamflow signatures. The general model structure is explained in Sect. ~~4.1~~3.2.1, the error model and the calibration procedure are described in Sect. ~~4.1~~3.2.2 and ~~4.1~~3.2.3, 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~~3.2.4.

~~4.1.1~~3.2.1 ~~4.1.1~~ General structure of the hydrological model

We describe here the general model structure~~. Specific choices for~~; the ~~various~~definition of specific model experiments~~ are~~, which depends on the results of the signatures analysis done in the first step, will be described in ~~Section~~Sect. 4.2.1.~~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~~is due to the necessity of having locations in the model where the streamflow ~~was~~is both observed and simulated and, therefore, it ~~was~~is 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, 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 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~~are tested, as described in Section 4.2.1.~~5~~.

Each HRU has a unique parameterization. However, ~~given the choice of discretizing~~depending on how the inputs ~~per subcatchment, a HRU that spans multiple subcatchments will generally~~are discretized, the same HRU can have different states in ~~each subcatchment.~~different parts of the catchment. Therefore, the same HRU needs its own model representation whenever the spatial variability of states needs to be considered. For example, if the inputs are discretized per subcatchment, the same HRU needs a separate 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~~In order to limit the ~~modelling framework SUPERFLEX (Fenicia et al., 2011). In contrast to Fenicia et al. (2016), for simplicity we chose a unique~~ levels of decisions of the semi-distributed models, some of the aspects of the distributed models are fixed a-priori, and others are left open. In particular:

- The structure chosen to represent the various HRUs ~~(as said above, this~~is kept fixed. That is, differences between HRUs will be reflected only through the parameter values.
- The definition of HRUs is left open. In particular, we do not a-priori specify which approach is used to discretize the landscape.
- The spatial discretization of the model inputs is left open. Hence, we do not decide in advance which spatial discretization of the inputs is most appropriate.

Only the fixed decision about the HRUs model structure ~~will generally have different parameters in order to represent the hydrological behaviour of distinct HRUs). The~~ is here described, whereas the open decisions are described in the Results section (Sect. 4.2.1). The spatial organization of the model structure ~~used to represent the HRUs is~~ is represented in Fig. 6 with the equations listed in the Appendix A. 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, proceeded by a lag function to offset the hydrograph, and a slow reservoir (SR), designed to represent baseflow. This structure 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. The model was built using the modelling framework SUPERFLEX (Fenicia et al., 2011).

~~4.1.2~~3.2.2 ~~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 $\boldsymbol{h}(\boldsymbol{\theta}_{\mathrm{h}},\boldsymbol{x})$ (illustrated in Sect. ~~4.1~~3.2.1) and a random residual error term $\boldsymbol{E}(\boldsymbol{\theta}_{\mathrm{E}})$ that accounts for all data and model uncertainties ($\boldsymbol{\theta}_{\mathrm{h}}$ and $\boldsymbol{\theta}_{\mathrm{E}}$ represent the hydrological and the error parameters):

$$z[\boldsymbol{Q}(\boldsymbol{\theta},\boldsymbol{x});\lambda] = z[\boldsymbol{h}(\boldsymbol{\theta}_{\mathrm{h}},\boldsymbol{x});\lambda] + \boldsymbol{E}(\boldsymbol{\theta}_{\mathrm{E}}) \tag{4}$$

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

$$z[y_{\mathrm{t}};\lambda] = \frac{y_{\mathrm{t}}^{\lambda}-1}{\lambda} \tag{5}$$

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

$$E_{\mathrm{t}} \sim N(0;\sigma^2) \tag{6}$$

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

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

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

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

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~~3.2.3.

~~4.1.3~~3.2.3 ~~4.1.3~~ Calibration

Parameter calibration ~~was~~is performed with the objective of maximizing their posterior 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~~is used for the parameters, this is equivalent to maximizing the likelihood function in the defined parameter space; the optimization procedure ~~was~~is performed with a multi–start quasi–Newton method (Kavetski et al., 2007) with 20 independent ~~searchers~~searches. 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~~is carried out in space–time validation (see also Fenicia et al., 2016). For this purpose, the time domain ~~was~~is 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~~are split 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 quadrants, such that the model can be calibrated in one quadrant and validated in the other three. For space–time validation, the model ~~was~~is 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~~is then combined, to calculate model performance using various indicators (see Sect. ~~4.1~~3.2.4). Results are presented for space time validation, which represents the most challenging test of model performance.

~~4.1.4~~3.2.4 ~~4.1.4~~ Performance assessment

Model performance ~~was~~is 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. (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

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

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) ~~which, as~~selected using the procedure illustrated in ~~Section~~Sect. 3.1.2.~~1, are average daily streamflow (ζ_{Q}), baseflow index (ζ_{BFI}) mean half streamflow date (ζ_{HFD}), 5$^{\mathrm{th}}$ percentile of the streamflow (ζ_{Q5}), and duration of high flow events (ζ_{HQD}).~~ The accordance between simulated and observed signatures ~~was~~is assessed both visually and using the Spearman's rank correlation.

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

4 Results and interpretation

4.1 Influencing factors on the spatial variability of streamflow signatures

This section illustrates the results of the correlation analysis complemented by expert judgement aimed to identify influencing factors that control the spatial variability of streamflow signatures; Section 4.1.1 presents the results of the selection of meaningful statistics; Section 4.1.2 identifies climate and landscape indices controlling streamflow signatures and presents consequences for model development; Section 4.1.3 formulates the hypotheses that have to be tested by the hydrological model.

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

The streamflow signatures defined in Sect. 34.1.1 were calculated for each subcatchment and the values are shown in Table 3 together with the coefficient of variation. All the signatures have a coefficient of variability 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, a subset of ζ can be defined as follows:

- 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 used in the subsequent analyses: average daily streamflow (ζ_{Q}), baseflow index (ζ_{BFI}), half streamflow period (ζ_{HFD}), 5th percentiles of the streamflow (ζ_{Q5}), and duration of high-flow events (ζ_{HQD}).

In terms of climatic indices, Table 4 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 %).

Figure 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 5 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.

4.1.2 Selection of controlling factors on streamflow signatures

Figure 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 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 sufficient to reach a p-value lower than 0.05.

These results are the premise for designing meaningful model experiments.

4.1.3 Hypotheses for model building

In this section, we synthetize the outcomes of previous analyses in the form of testable hypotheses for model building.

1. The precipitation is the first driver of the differences in the water balance of the subcatchments. The effect of topographic variability manifests itself primarily as an influence on precipitation (amount and type). Accounting for variability of precipitation therefore implicitly reflects such effect of topography on the hydrograph, since some inputs were interpolated taking into account the effect of the elevation (Sect. 2). Other phenomena potentially altering the water balance (e.g. regional groundwater flow) do not have a significant role and should not be considered.
2. Snow related processes (e.g. amount of snow, timing of snowmelt) control differences in streamflow seasonality between subcatchments. Hence, the model needs to account for snow related processes and their spatial variability.
3. Geology exerts an important control on the partitioning between quick flow and baseflow. Hence, the model should distinguish the different response behaviour of distinct geological areas.
4. The other catchment characteristics (e.g. soil, vegetation) show little or no correlations with the streamflow signatures and therefore they should not be considered if the idea is to keep the model as simple as possible.

These hypotheses will be tested through specific model comparisons, described in Sect. 4.2.1.

4.2 Modelling

This section presents the results of the modelling experiments. ~~5~~ Section 4.2.1 describes the model comparisons designed to test the hypotheses formulated in section 4.1.3. Section 4.2.2 illustrates model results in terms of hydrograph metrics. Section 4.2.3 presents model results in terms of signatures. An interpretation of the results, including a comparison with the conclusions of the signatures analysis, is given in Sect. 4.2.4.

~~4.1.5~~4.2.1 Model experiments for testing the results of the ~~correlation~~signatures analysis

Using the model structure described in Sect. ~~4.1~~3.2.1, four model configurations were compared by varying the number and the definition of the HRUs, and changing the structure of the HRUs (Fig. 6). The objective of the experiments was to test the hypotheses 1-4 in Sect. 4.1.3~~.3~~ using semi-distributed hydrological models.

~~The~~For all models, the meteorological inputs (precipitation, PET, temperature) are aggregated at the subcatchment scale. Based on the first hypothesis ~~(precipitation controls the~~ in Section 4.1.3, we assume that this discretization is sufficient to capture the regional difference in water balance~~)~~ between subcatchments. This hypothesis is tested with the model M0, with uniform parameters ~~on~~in 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 to reproduce differences in streamflow averages between subcatchments.

The second hypothesis in Section 4.1.3 (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. We expect that this model will be able to reproduce differences in streamflow seasonality between subcatchments.

The third hypothesis in Section 4.1.3 (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). We expect that M2 will be able to reproduce differences in the baseflow index between subcatchments.

The fourth hypothesis in Section 4.1.3 (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~~in terms of complexity but the 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. This model is as complex as M2 (therefore it is more complex than M1); hence it has the same dimensions of flexibility to fit the data. However, since the structure of this model does not incorporate the cause-effect relationships derived from the signatures analysis, we expect that its predictive performance will be poorer than M2.

The total number of the calibrated parameters depends on the number of HRUs and on the structure used to represent them: it was 8 for M0, 9 in M1, and 13 in M2 and M3, where 5 parameters were linked between different HRUs (Fig. 6 and Table A1); those parameters are: C_{e} that governs the evapotranspiration, $t_{\mathrm{rise}}^{\mathrm{OL}}$ and $t_{\mathrm{rise}}^{\mathrm{IL}}$ that control the routing in the river network, k_{WR} that regulates the outflow of the snow reservoir, and $S_{\mathrm{max}}^{\mathrm{UR}}$ that determines the behaviour of the unsaturated reservoir.

~~4.2 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 correlation analysis, is given in Sect. 4.2.3.~~

~~4.2.1~~4.2.2 ~~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 four models in calibration and validation. It can be observed that M0 is, by far, the worst model, ~~having a low~~with the lowest value of the likelihood~~. Moving to~~ function. Regarding the other three models, it can be seen that, during calibration, M1, which has the lowest number of calibration parameters, has the lowest performance, whereas M2 and M3 have higher and similar ~~higher~~ likelihood values. This behaviour 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, ranking, respectively, second and the first in terms of optimal likelihood value.

The likelihood function represents an aggregate ~~metric~~metrics 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), M3 (NS = 0.77), and M0 (NS = 0.68). M3 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~~in Andelfingen, Frauenfeld and Wängi, and clearly worse ~~on~~in 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~~mostly 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~~4.2.3 ~~4.2.2~~ Model performance in terms of signature metrics

Figure 8 compares the ability of M0 and M1 to capture the signatures representing average streamflow (ζ_{Q}) and seasonality (ζ_{HFD}). The analysis is presented for space–time validation 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 represents where the dots should align in case of perfect simulation results. The Spearman's rank score is also reported and gives information about the degree of dependency between the two variables. It is important to stress that the models have not been calibrated using any of the signatures as objective function, which therefore represent an independent evaluation ~~metric~~metrics.

It can be observed that M0 represents ζ_{Q} as well as M1, with almost no difference between the two models. Focusing on the ability of capturing~~-~~ ζ_{HFD}, it can be seen that ~~with M0~~ the points corresponding to M0 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 0.66 for M0 and 0.85 for M1.

Figure 9 compares the observed and simulated signatures for the other three models (M1, M2, and 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) 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~~are still aligned ~~to~~along the diagonal but ~~it is~~are quite dispersed, especially if compared with ζ_{Q} and ζ_{HFD}, meaning that the models capture the general tendency but have deficiencies capturing the inter-annual variability.

In terms of ζ_{BFI}, M2 performs clearly better than the other models. It is the only model 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~~are quite dispersed and the dots align almost vertically, implying that the simulated values have a range of variability that is definitely smaller than the observed data.

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~~as an 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~~4.2.4 ~~4.2.3~~ Interpretation of hydrological model results

The results of the hydrological model experiments appear to support our hypothesis that only models that account for the influencing factors that affect the streamflow signatures are able to reproduce streamflow spatial variability (see Sect. ~~3~~4.1.3). This provides confidence that those models are a realistic representation of dominant processes in the catchment.

In particular, the results of M0 show that accounting for the spatial heterogeneity of the precipitation alone is sufficient to achieve a good accuracy signatures of water balance, with r of 0.95 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 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} goes from 0.31 for M1 to 0.83 for M2. This result confirms that geology influences spatial variability of quickflow vs baseflow partitioning, as indicated by ~~correlation~~signatures 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 ζ_{BFI}, M2 is still much better (e.g. the Spearman's rank score for ζ_{BFI} is 0.83 for M2 and 0.52 for M3).

All the models do not preform particularly well in reproducing ζ_{Q5} and ζ_{HQF}. These problems ~~shows~~show that such models may not represent well extreme values (high and low ~~flow~~flows), and therefore they are still amenable to improvements. Overall, distributing the inputs in space and accounting for the spatial distribution of snow related processes is sufficient to get good performance metrics~~,~~ of water balance~~,~~ and seasonality, confirming the fact that only the precipitation rate and the partitioning between rainfall and snow are the first controllers on these hydrograph characteristics~~, but, if we want~~. However, in order to capture ~~also~~ other important characteristics of the hydrograph, described by 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. 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). This means that 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.

5 ~~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 correlation based analyses, which seek correlations between 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 correlation analysis and expert judgement for identifying dominant influencing factors on streamflow signatures with hydrological modelling, by using the interpretation of the correlation analysis as an inspiration for generating testable model hypotheses. The combination of correlation analysis on streamflow signatures and hydrological modelling is beneficial because on the one hand, the speculations on dominant processes resulting from the 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 ~~correlation~~signatures 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 ~~correlation~~signatures analysis can motivate some of these decisions. For example, the HRUs defined based on geology, as suggested by ~~correlation~~signatures 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~~4.1.2 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~~reasons, 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.~~ necessarily subjective. Although subjectivity is difficult to avoid, it is important being transparent about the decision taken and the argumentations on which they are based, how weak or strong they may be, so that they can be reappraised and revised if new evidence is acquired.

The choice of streamflow signatures is based on the large-sample study from Addor et al. (2017), which provides a broad range of signatures typically used in hydrology. Our analyses showed that this selection is rather inclusive, with several strongly correlated signatures (e.g. ζ_Q and ζ_{RR}). For this reason, we eventually used a much smaller selection of the original set of signatures (12 in the original set vs. 5 in the final set). The apparent inclusivity of the set from Addor et al. (2017) provides confidence that the main properties of streamflow are captured in our study. However, it does not guarantee that this set of signatures is sufficient in representing streamflow time series.

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 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.~~ are still a small sample of the characteristics found elsewhere. Moreover, although the model evaluation uses validation in space and time, which is a relatively incisive test, the spatial validation is carried out in a nested setup. The application of systematic model development strategies to other places and scales, and spatial validation to entirely different regions, are necessary to obtain more generalizable insights.

The ~~limited~~small number of ~~catchments~~subcatchments involved in this study (~~only~~ 10) ~~can also pose some problems in~~ limits the range of viable methods for identifying relationships between landscape and climatic indices and streamflow signatures (Sect. 3.1) to rather simple approaches. In particular, our correlation analysis, ~~where only linear or~~ although accounting for non-linearity, is limited to monotonic correlations ~~have been investigated while~~between variables, and it is unable to identify other forms of relationship, including the mutual interaction between various influencing factors~~, have been neglected. This can lead to the exclusion of characteristics that~~. The usage of more advanced techniques, including machine learning approaches such as random forest or clustering analyses, are ~~indirectly related to the streamflow signatures~~most efficient when larger samples are available and could represent a more suitable choice in these situations.

6 ~~6~~ Conclusion

In this study, we presented a methodology for the construction of a semi-distributed hydrological model where model hypotheses, instead of being made a–priori, are informed by preliminary ~~correlation~~ analysis on determining the dominant climatic and landscape controls on streamflow ~~signatures~~spatial variability. 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 ~~correlation~~such preliminary analysis.

Our analysis was applied to the Thur catchment, ~~with the objective of understanding the main controls~~subdivided in 10 subcatchments based on ~~streamflow spatial variability~~available stream gauging stations. The main findings are summarized in the following points:

1. we found large spatial variability between the subcatchments of the Thur in terms of various 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~~among catchments. Other precipitation characteristics such as season, frequency and duration of dry and wet days did not vary significantly ~~between~~among 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. based on correlation analysis and expert judgment, we determined that climatic variables, especially precipitation average, are the main controls on streamflow average yearly values; 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;
3. the results of the ~~correlation~~signatures 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.~~accounting for the heterogeneity of precipitation, snow related processes, and landscape features such as geology, is necessary to produce hydrographs that have signatures similar to the observed ones. In particular, we confirmed that M0, in spite of a generally poor performance, is sufficient to capture signatures of streamflow average~~.~~, showing that only distributing the meteorological inputs is sufficient to explain regional differences in average streamflow and that other phenomena potentially altering the water balance (e.g. regional groundwater flows) do not play a significant role. M1 improves signatures of streamflow seasonality~~.~~, showing that snow is the main responsible for the variability of the seasonality among the catchments. M2 enables reproducing signatures such as the baseflow index, showing that incorporating the geology of the catchment is important for reproducing regional differences in baseflow. 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~~ leading to the same complexity~~, not only do not lead to an improvement, but~~ as M2, cause deterioration in model performance in space-time validation. Overall, these results confirm the hypotheses based on the signatures analysis and suggest that the 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.

7 Appendix

7.1 Appendix A: Hydrological model details

7.1.1 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.

8 Team list

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

9 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.

10 Competing interests

The authors declare that they have no conflict of interest.

11 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. Conrad Jackisch (editor), Dr. Lieke Melsen, Dr. Shervan Gharari, and the two anonymous ~~referee~~referees for their feedback and their help in improving this paper. MZ and MS dedicate this work ~~on the~~in memory of Dr. Joachim Gurtz, who pioneered the work on distributed hydrological modelling in the Thur catchment (Gurtz et al., 1999).

12 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.~~

Ajami, K. N., Gupta, H., Wagener, T., and Sorooshian, S.: Calibration of a semi-distributed hydrologic model for streamflow estimation along a river system, J Hydrol, 298, 112-135, https://doi.org/10.1016/j.jhydrol.2004.03.033, 2004.

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~~Antonetti, M., and Zappa, M.: How can expert knowledge increase the realism of conceptual hydrological models? A case study based on the concept of dominant runoff process in the Swiss Pre-Alps, Hydrol. Earth Syst. Sci., 22, 4425-4447, 10.5194/hess-22-4425-2018, 2018.

Arnold, J. G., Srinivasan, R., ~~Verderio, P., and Marubini, E.: Bravais-Pearson~~Muttiah, R. S., and ~~Spearman Correlation Coefficients: Meaning, Test of Hypothesis and Confidence Interval, The International~~Williams, J. R.: LARGE AREA HYDROLOGIC MODELING AND ASSESSMENT PART I: MODEL DEVELOPMENT1, JAWRA 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~~the American Water ~~Resour As, 40, 503-522, DOI~~Resources Association, 34, 73-89, 10.1111/j.1752-1688.~~2004.tb01046~~1998.tb05961.x, ~~2004~~1998.

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.

Boyle, D. P., Gupta, H. V., Sorooshian, S., Koren, V., Zhang, Z., and Smith, M.: Toward improved streamflow forecasts: value of semidistributed modeling, Water Resour Res, 37, 2749-2759, 10.1029/2000wr000207, 2001.

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.

Gao, H., Hrachowitz, M., Fenicia, F., Gharari, S., and Savenije, H. H. G.: Testing the realism of a topography-driven model (FLEX-Topo) in the nested catchments of the Upper Heihe, China, Hydrol. Earth Syst. Sci., 18, 1895-1915, 10.5194/hess-18-1895-2014, 2014.

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.

McInerney, D., Thyer, M., Kavetski, D., Githui, F., Thayalakumaran, T., Liu, M., and Kuczera, G.: The Importance of Spatiotemporal Variability in Irrigation Inputs for Hydrological Modeling of Irrigated Catchments, Water Resour Res, 54, 6792-6821, 10.1029/2017wr022049, 2018.

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.

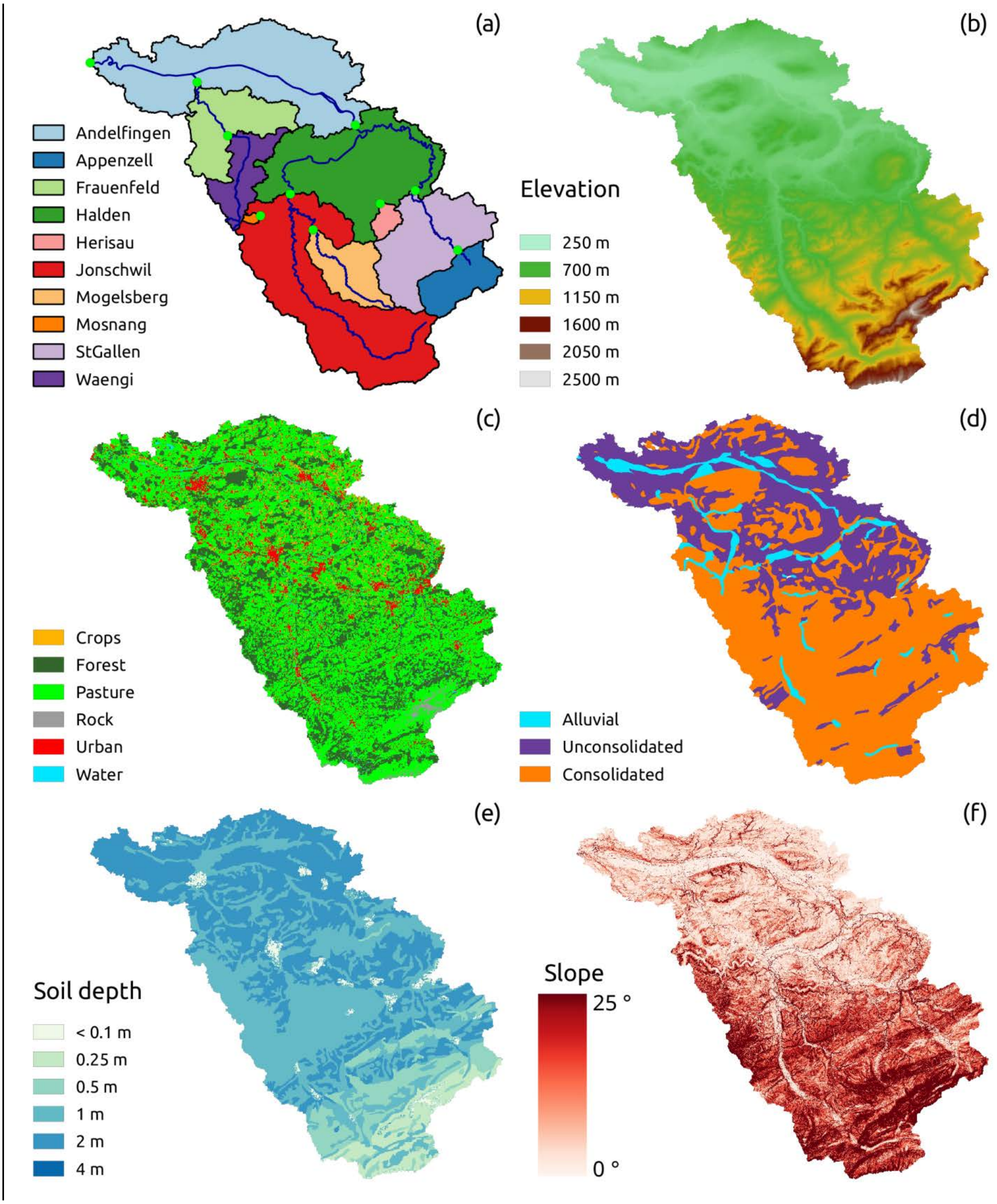
von Freyberg, J., Moeck, C., and Schirmer, M.: Estimation of groundwater recharge and drought severity with varying model complexity, J Hydrol, 527, 844-857, 10.1016/j.jhydrol.2015.05.025, 2015.

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.

Wooldridge, S., Kalma, J., and Kuczera, G.: Parameterisation of a simple semi-distributed model for assessing the impact of land-use on hydrologic response, J Hydrol, 254, 16-32, https://doi.org/10.1016/S0022-1694(01)00489-9, 2001.

13 Figures


(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 °

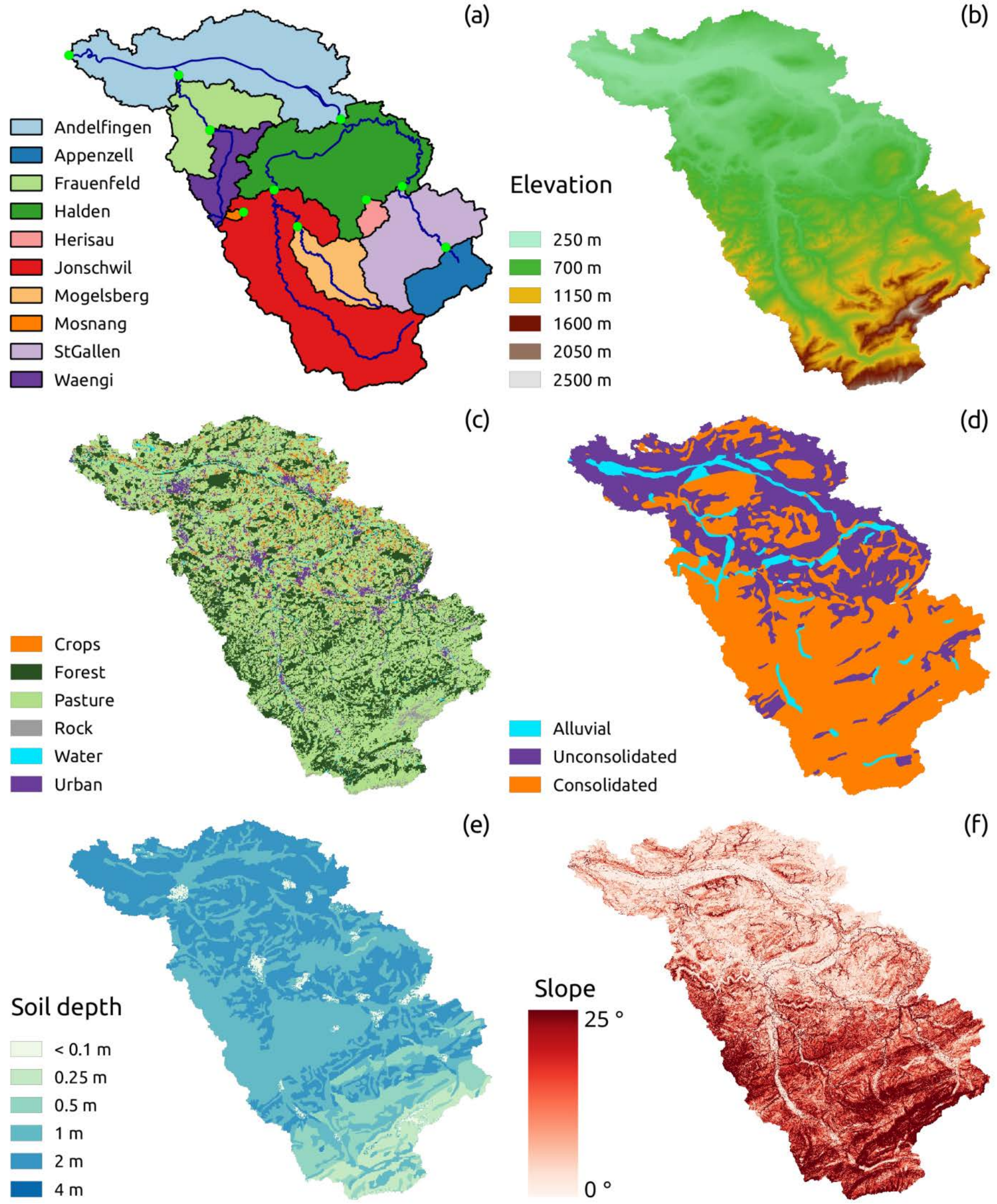


Figure 1: Landscape characteristics of the Thur catchment: (a) subdivision in subcatchments, river network, and gauging stations; (b) elevation map; (c) land use map; (d) simplified geology map; (e) soil depth map; (f) slope map (derived from the elevation map).

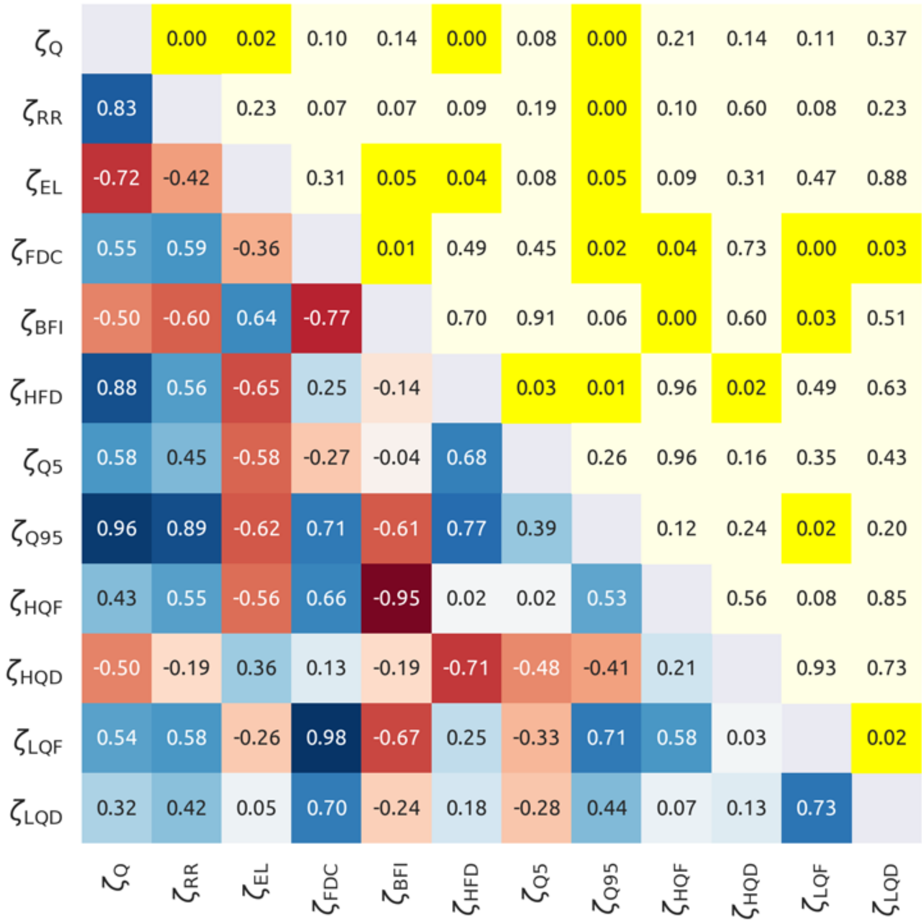


Figure 2: Internal correlation between the streamflow signatures. The lower triangle shows the Spearman's rank score with the red colour that indicates negative correlations and the blue that indicates positive correlations. The upper triangle reports the corresponding p-values, where yellow colour indicates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

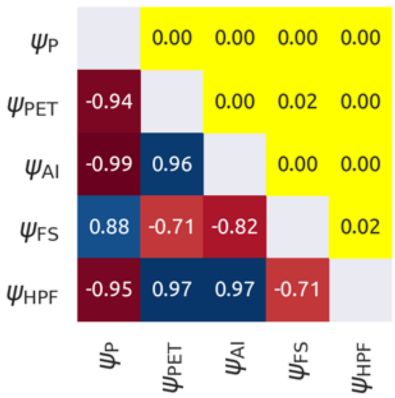


Figure 3: Internal correlation between the climatic indices. The lower triangle shows the Spearman's rank score with the red colour that indicates negative correlations and the blue that indicates positive correlations. The upper triangle reports the corresponding p-values, where yellow colour indicates a statistically significant correlation (p-value < 0.05). The symbols used in the figure are reported in ~~Sect. 3.1.1~~Table 2.

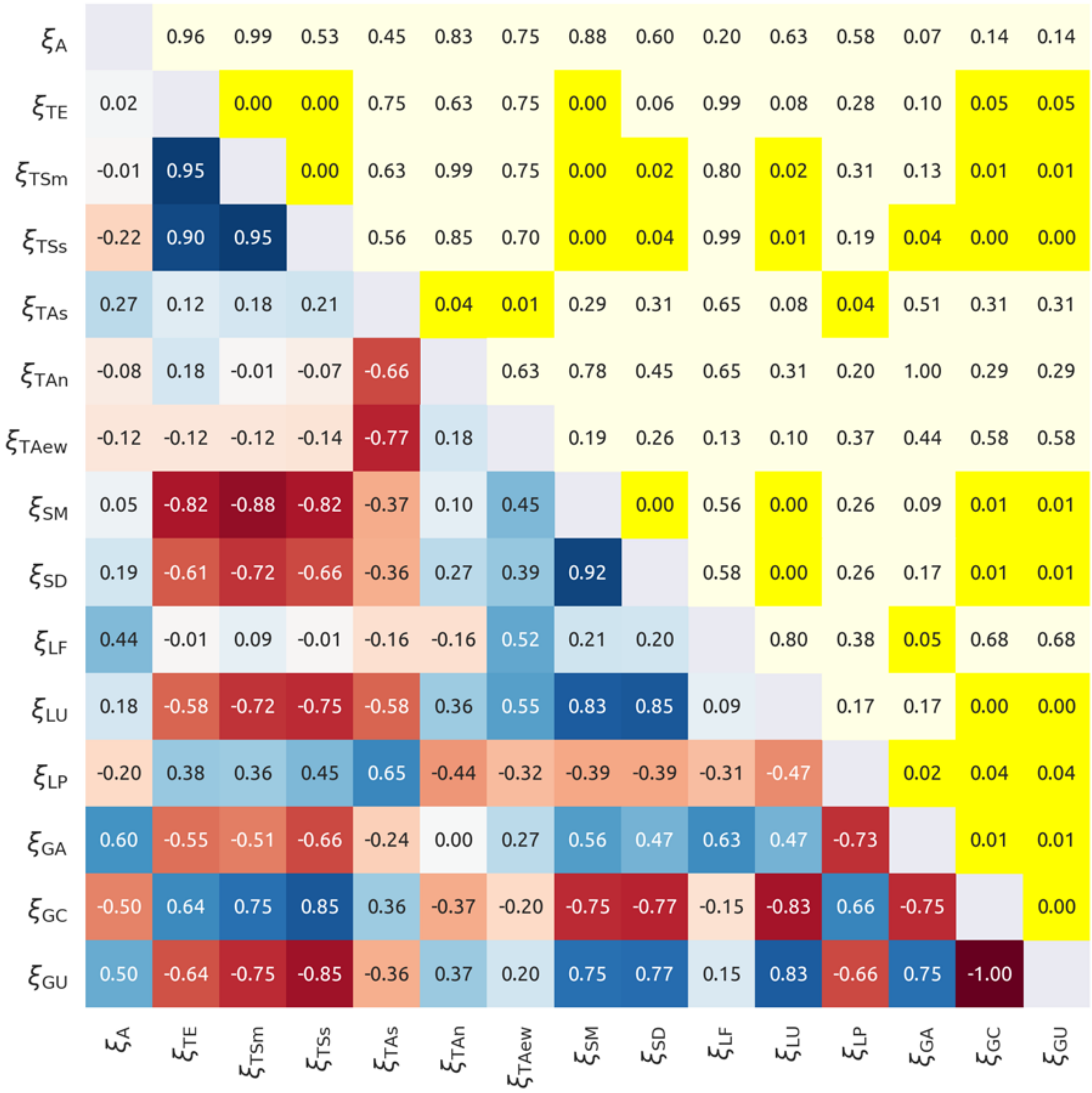


Figure 4: Internal correlation between the catchment characteristics. The lower triangle shows the Spearman's rank score with the red colour that indicates negative correlations and the blue that indicates positive correlations. The upper triangle reports the corresponding p-values, where yellow colour indicates a statistically significant correlation (p-value < 0.05). The symbols used in the figure are reported in ~~Sect. 3.1.1~~Table 2.

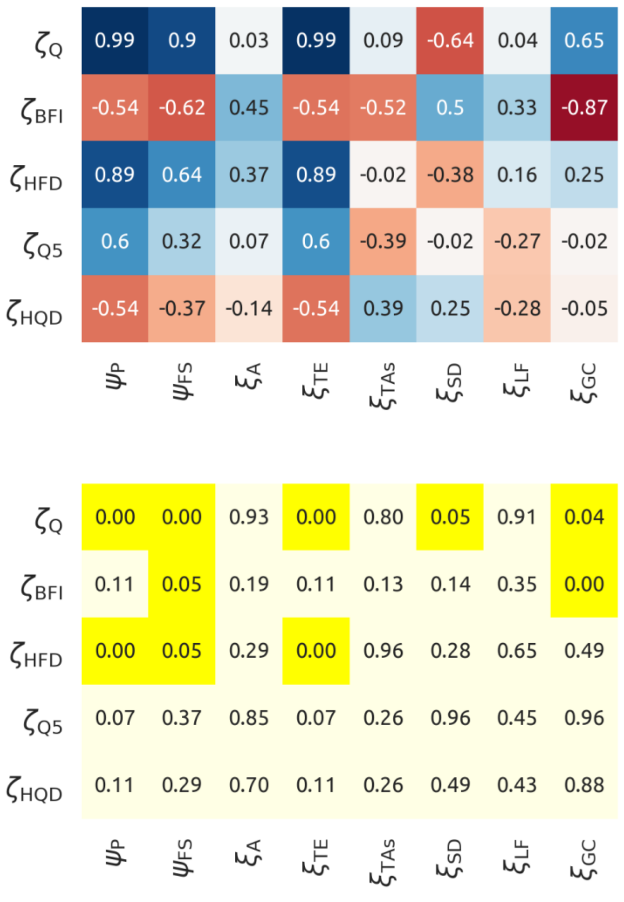
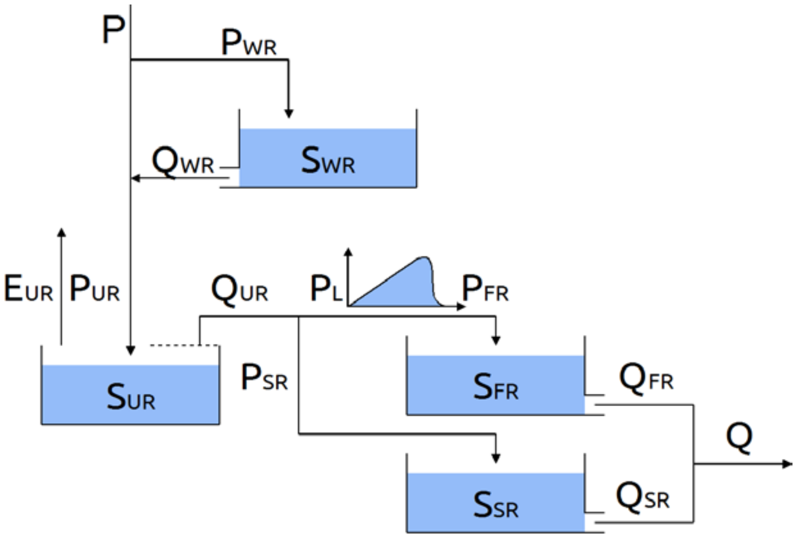
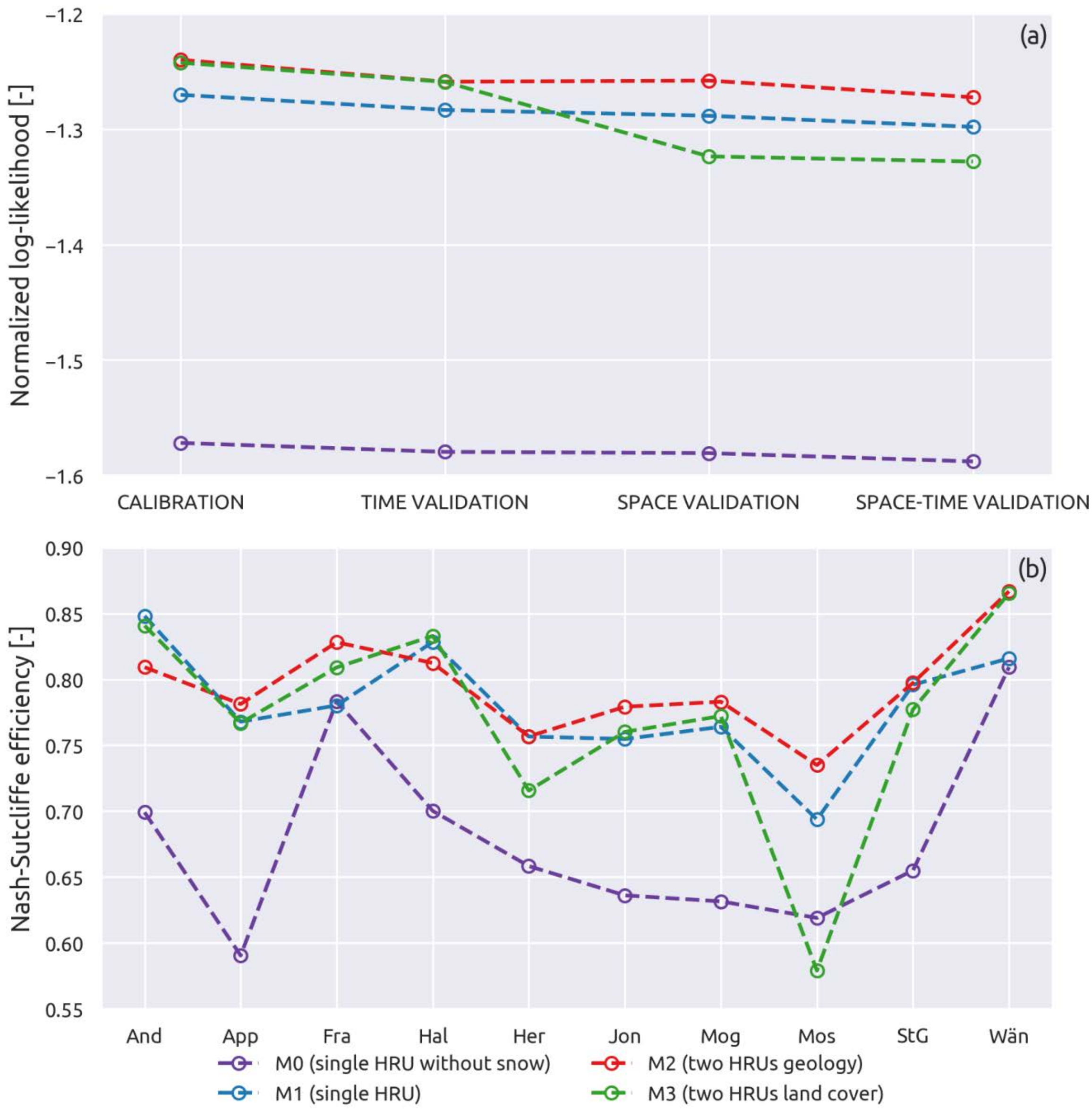


Figure 5: Correlation between the selected streamflow signatures (rows) and the selected climatic indices and catchment characteristics (columns). The upper panel shows the Spearman's rank score with the red colour that indicates negative correlations and the blue that indicates positive correlations. 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~~Table 2.



~~**Figure 6: Schematic representation of the model structure used for the HRUs in all the model configurations. 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**~~



−1.2
−1.3
−1.4
−1.5
−1.6
(a)
Normalized log-likelihood [-]
CALIBRATION
TIME VALIDATION
SPACE VALIDATION
SPACE-TIME VALIDATION
0.90
0.85
0.80
0.75
0.70
0.65
0.60
0.55
(b)
Nash-Sutcliffe efficiency [-]
And
App
Fra
Hal
Her
Jon
Mog
Mos
StG
Wän
M0 (single HRU without snow)
M1 (single HRU)
M2 (two HRUs geology)
M3 (two HRUs land cover)

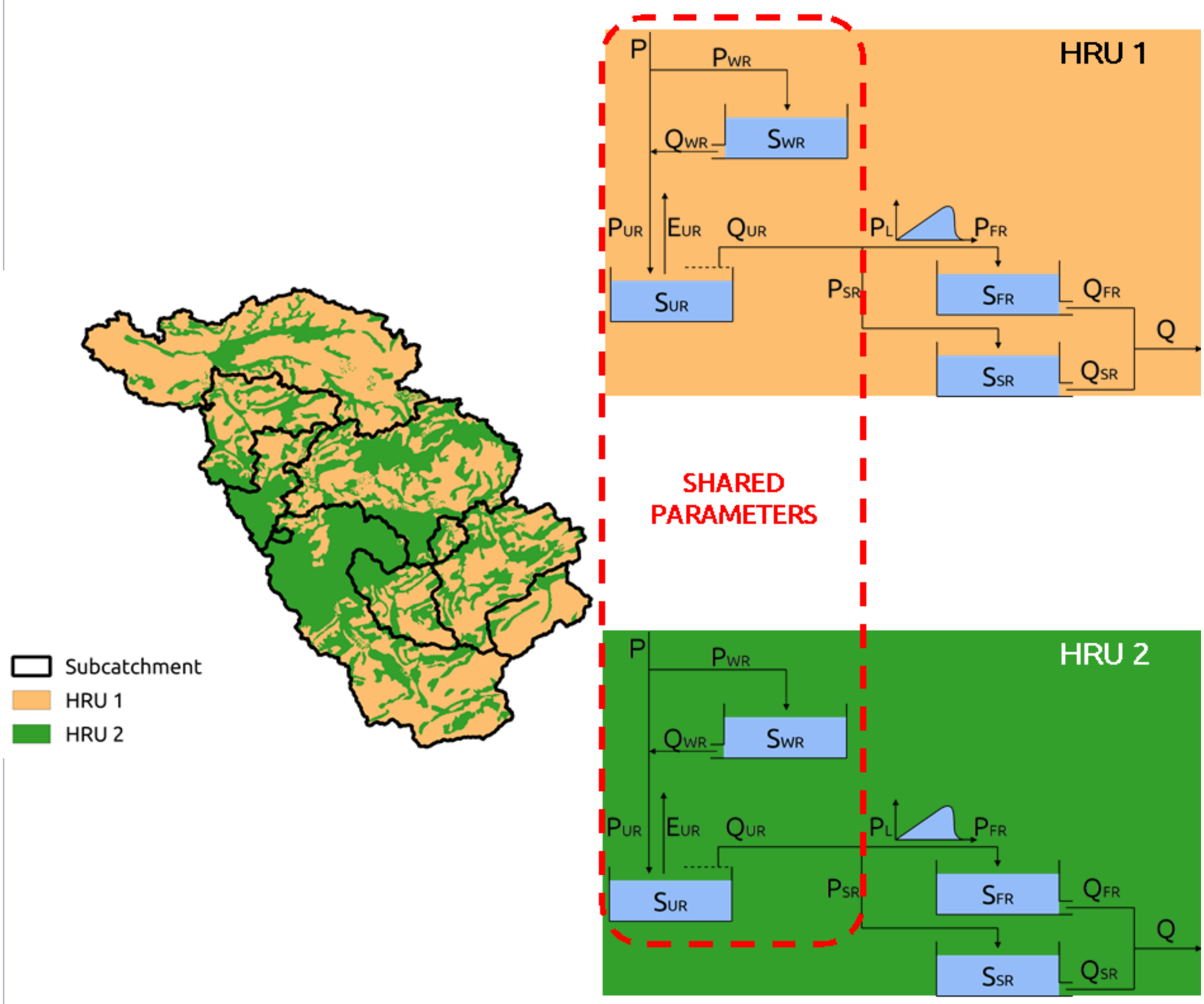


Figure 6: Spatial organization of the model structure: the catchment is divided in subcatchments (black lines), based on the location of the gauging stations, and HRUs (background colour), based on the catchment characteristics. All the HRUs have the same structure but each HRU has its own parameterization except for some shared parameters. In the case of a single HRU model (i.e. M0 and M1), the model maintains the subdivision in subcatchments but loses the subdivision in multiple HRUs.

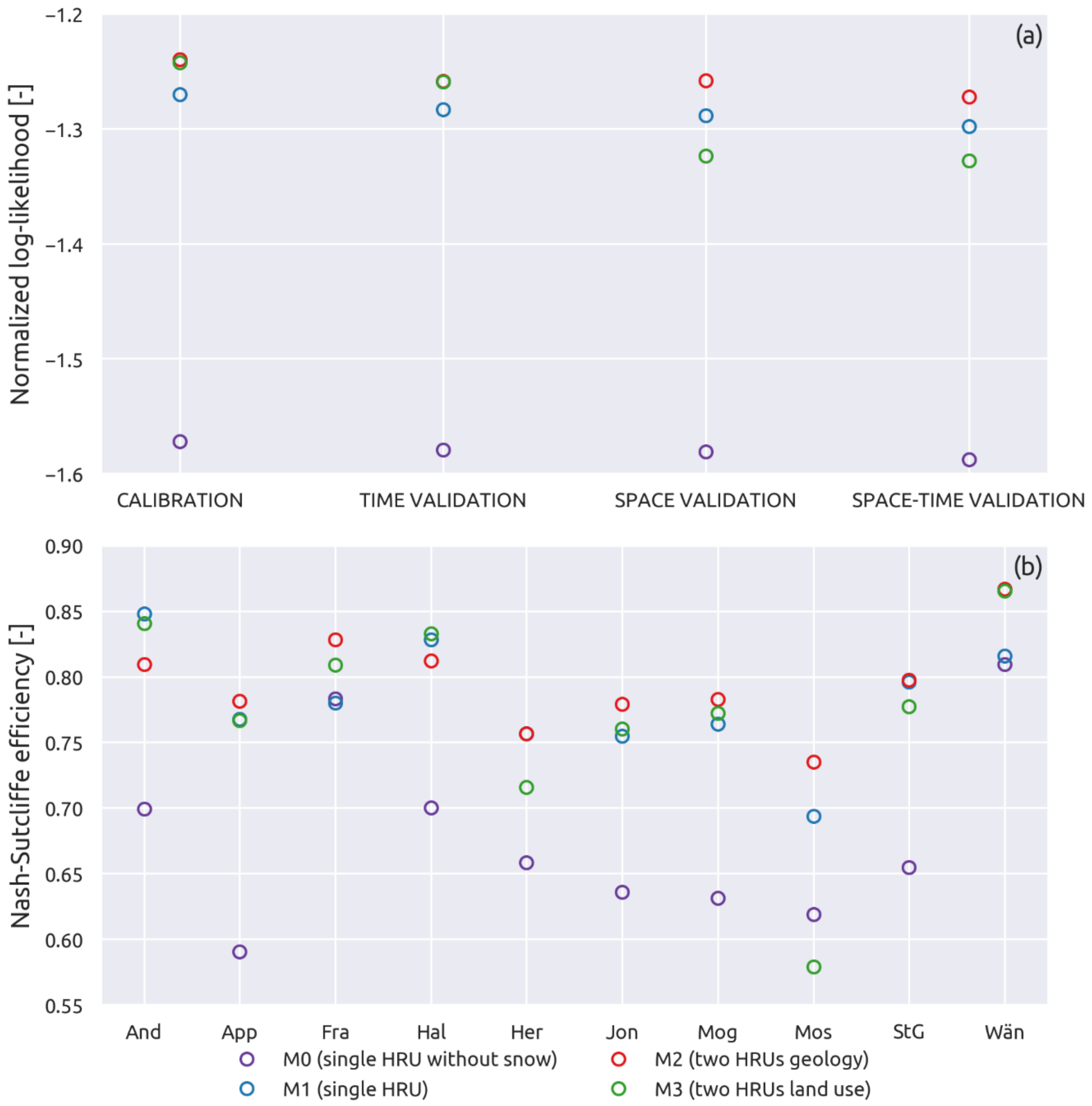


Figure 7: Normalized log–likelihood (a) and Nash–Sutcliffe efficiency (b) for the three model configurations. The upper plot (a) reports the variation between calibration and validation of the average of the 10 subcatchments; the lower plot (b) shows the variation between subcatchments during space–time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

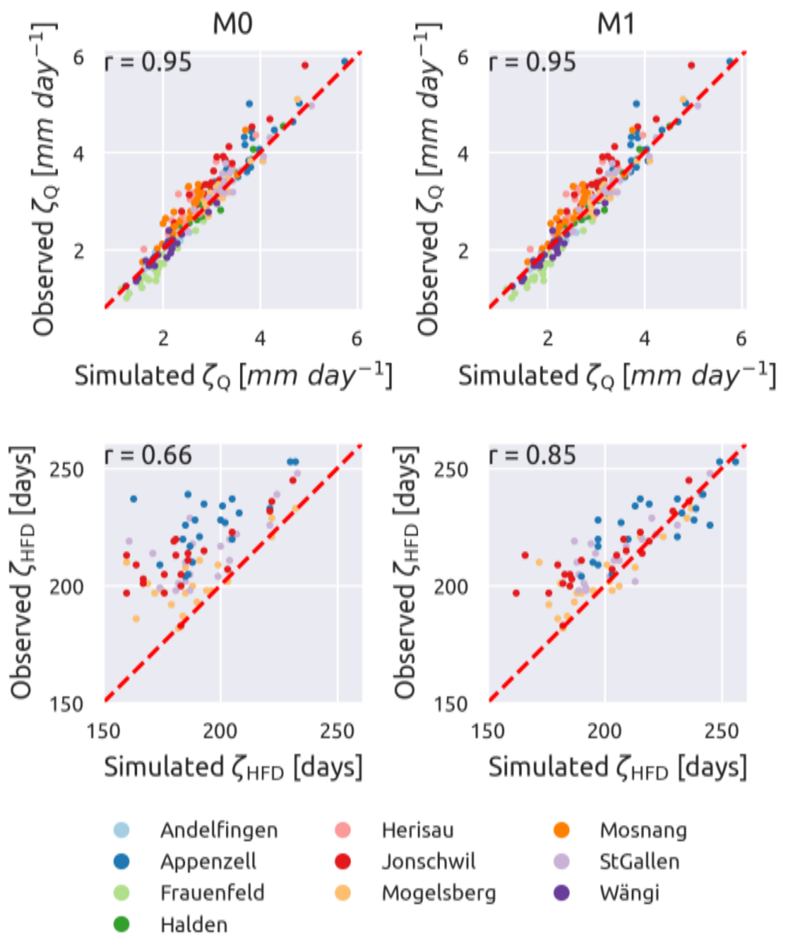


Figure 8: Influence of the model structure on the representation of the average streamflow (ζ_Q) and the mean half streamflow day (ζ_{HFD}). Single HRU model without snow reservoir on the left~~,~~ (M0), single HRU model with snow reservoir on the right~~.~~ (M1). Each dot represents a year and each 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 indicates where all points should align in case of perfect match. The Spearman's rank score (r) is also reported.

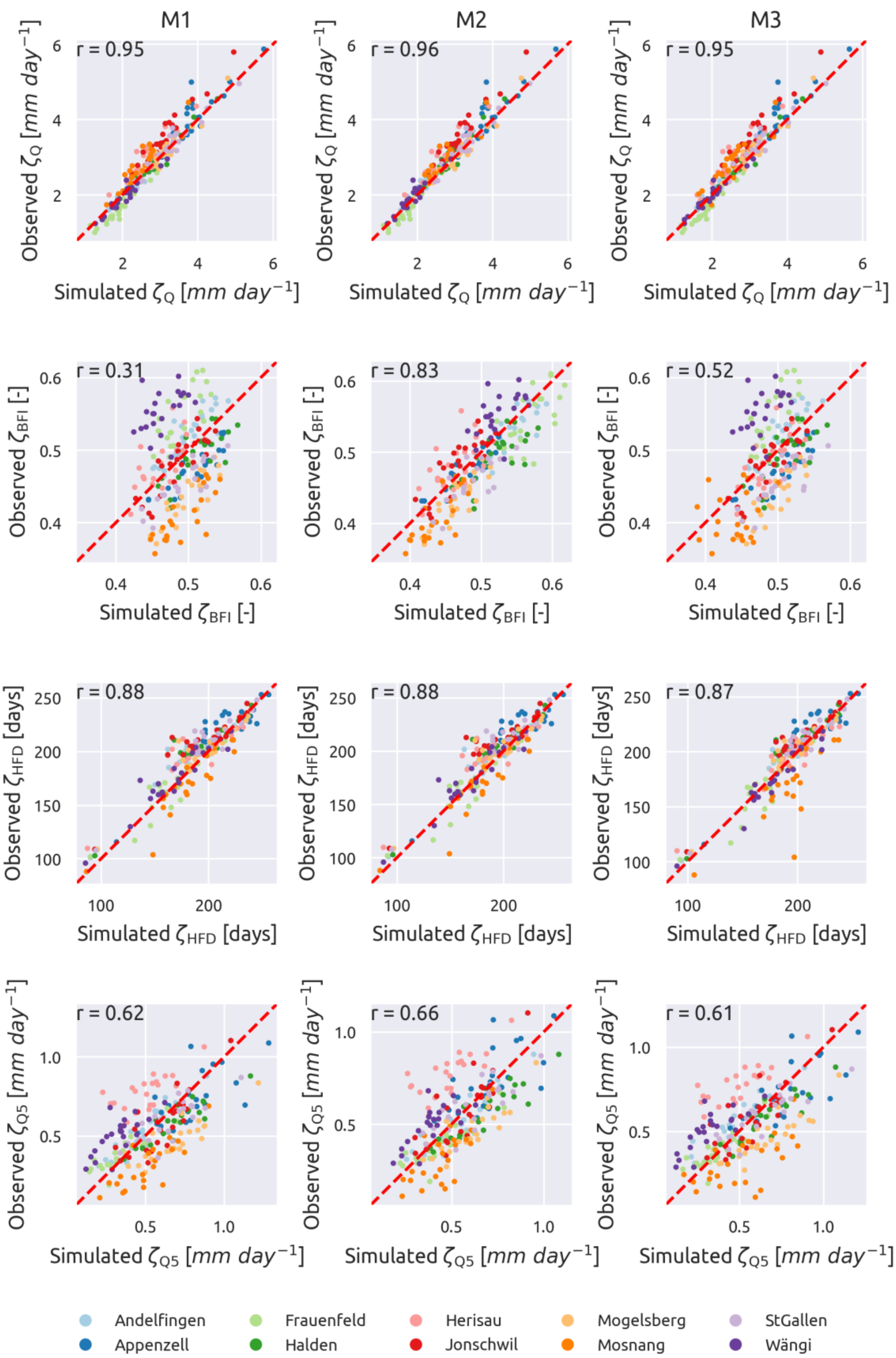


Figure 9: Simulated vs observed streamflow signatures. Single HRU model on the left~~,~~ (M1), two HRUs model based on geology in the centre~~,~~ (M2), two HRUs model based on land use on the right~~.~~ (M3). Each dot represents a year and each colour a subcatchment. From up to bottom, mean daily streamflow (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.

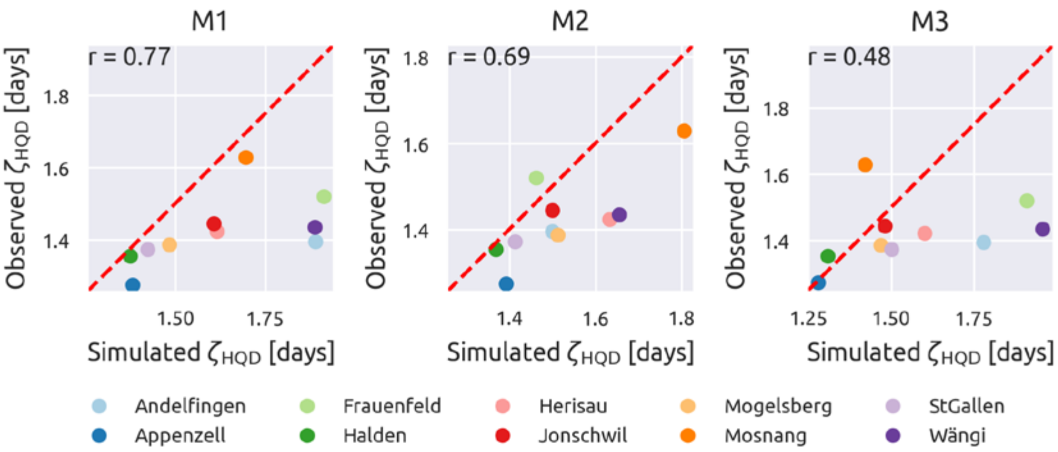


Figure 10: Ability of the hydrological models of representing the signature duration of low-flow events (ζ_{HQD}). Single HRU model on the left, two HRUs model based on geology in the centre, two HRUs model based on land use on the right.

14 Tables

Table 1: Identification of the gauging stations and description of the 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: List of streamflow signatures, climatic indices, and subcatchments characteristics considered in the study.

Symbol	Name
Streamflow signatures	
ζ_{Q}	Average daily streamflow
ζ_{RR}	Runoff ratio
ζ_{EL}	Streamflow elasticity
ζ_{FDC}	Slope of the flow duration curve
ζ_{BFI}	Baseflow index
ζ_{HDF}	Mean half streamflow date
ζ_{Q5}	5^{th} percentile of the streamflow
ζ_{Q95}	95^{th} percentile of the streamflow
ζ_{HQF}	Frequency of high-flow events
ζ_{HQD}	Mean duration of high-flow events
ζ_{LQF}	Frequency of low-flow events
ζ_{LQD}	Mean duration of low-flow events
Climatic indices	
ψ_{P}	Average daily precipitation
ψ_{PET}	Average daily potential evapotranspiration
ψ_{AI}	Aridity index
ψ_{FS}	Fraction of snow
ψ_{HPF}	Frequency of high-precipitation events
ψ_{HPD}	Mean duration of high-precipitation events
ψ_{HDS}	Season with most high-precipitation events
ψ_{LPF}	Frequency of low-precipitation events
ψ_{LPD}	Mean duration of low-precipitation events
ψ_{LPS}	Season with most low-precipitation events
Subcatchments characteristics	
ξ_{A}	Subcatchment area
ξ_{TE}	Average elevation
ξ_{TSm}	Average slope
ξ_{TSs}	Fraction of the subcatchment with steep areas
ξ_{TAs}	Fraction of the subcatchment facing south
ξ_{TAn}	Fraction of the subcatchment facing north
ξ_{TAew}	Fraction of the subcatchment facing east or west
ξ_{SM}	Average soil depth
ξ_{SD}	Fraction of the subcatchment with deep soil
ξ_{LF}	Fraction of the subcatchment with forest land use
ξ_{LC}	Fraction of the subcatchment with crops land use
ξ_{LU}	Fraction of the subcatchment with urbanized land use

ξ_{LP}	Fraction of the subcatchment with pasture land use
ξ_{GA}	Fraction of the subcatchment with alluvial geology
ξ_{GC}	Fraction of the subcatchment with consolidated geology
ξ_{GU}	Fraction of the subcatchment with unconsolidated geology

~~Table 2~~

Table 3: Values of the streamflow signatures. The names of the subcatchments are abbreviated using the first three letters~~.~~, the symbols are reported in Table 2. 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~~4: Values of the climatic indices. The names of the subcatchments are abbreviated using the first three letters~~.~~, the symbols are reported in Table 2. 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~~5: Values of the subcatchment characteristics. The names of the subcatchments are abbreviated using the first three letters~~.~~, the symbols are reported in Table 2. 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 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}	$\mathrm{d}^{-1}\mathrm{mm}^{-2}$	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 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)	$Q_{\mathrm{WR}} = M_{\max}^{\mathrm{WR}} f_{\mathrm{e}}(S_{\mathrm{WR}} \mid 2)$
Unsaturated reservoir (UR)	$\overline{S_{\mathrm{UR}}} = \frac{S_{\mathrm{UR}}}{S_{\max}^{\mathrm{UR}}}$
Unsaturated reservoir (UR)	$E_{\mathrm{UR}} = C_{\mathrm{e}}(PET) f_{\mathrm{m}}(S_{\mathrm{UR}} \mid 0.01)$
Unsaturated reservoir (UR)	$Q_{\mathrm{UR}} = P_{\mathrm{UR}} f_{\mathrm{p}}(\overline{S_{\mathrm{UR}}} \mid 2)$
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 $f_{\mathrm{e}}(x\|\theta) \to 1$ as $x \to \infty$
$f_{\mathrm{p}}(x\|\theta) = x^{\theta}$	Power function
$f_{\mathrm{m}}(x\|\theta) = \dfrac{x(1+\theta)}{x+\theta}$	Monod–type kinetics, adjusted so that $f_{\mathrm{m}}(1\|\theta) = 1$