Response to referee 2

To facilitate the review process, all referee comments are in **black** while the authors response in **blue**.

This paper provides an interesting assessment of geological information on streamflow signatures. The study of this work at separate scales is potentially interesting and a refreshing perspective for large-sample hydrology studies. The study could be a useful addition to the hydrological literature, but before I can recommend publication, the following points need to be clarified:

We sincerely thank the reviewer for their time to review our manuscript. We greatly appreciate their thoughtful and constructive feedback, which we found to be both valuable and encouraging. We have carefully considered each comment and have provided detailed responses to all of them. We hope that our revisions and clarifications can bring the paper closer to being suitable for publication.

Please note that, due to their closely related nature, the next four comments are addressed together in a combined response.

- "Normal" (typical) large sample hydrology analyses study (statistics) relationships across the entire study domain at once. This paper takes a different approach by studying local (basin-scale) relationships across many basins at once. This is a valuable addition to the literature, but this contrast needs to be better highlighted, as it fundamentally changes the question that is asked (the current paper does not check continental scale dominance of catchment attributes) but the local scale dominance of factors (across an entire continent). This different approach inherently changes the answers one gets, but will also change the question that is asked, and as such this should be better contrasted to the existing literature. Right now this
- In the introduction, the manuscript reflects on why landscape appears to have a small impact on hydrology according to LSH studies. The work states this may be due to several reasons but maybe overlooks (according to my gut feeling, not that I have no formal evidence) the most obvious reason: the landscape can have a very important role on hydrology but how important this role is depends on the diversity of climates considered. Most LSH hydrology studies are across tremendous climate gradients (e.g. the current continental-scale study) and this easily dominates the role of many (but not all) hydrological signatures. Sure, landscape will (strongly) modify how incoming precipitation is partitioned, but, if climate differences are big enough, it cannot override these climate effects. And if we are not yet super successful at normalizing for climate effects, effectively learning about landscape in LSH studies across large geographical domains becomes challenging. If LSH were conducted with many catchments in a similar climate setting, the role of landscape would become (relatively seen) much more important and very likely easier (but still challenging, due to reasons you also state) to identify.
- The approach that is used already seems to partly acknowledge this, as at large scales, it checks for correlations at the scale of a basin, which reduces climate gradients. Explain this. But also explain thereby that the results you get are "dominant controls" on the scale that you study, and not at the continental scale when the large scales is studied.
- In the discussion, the paper also reflects on this by comparing it to previous studies "(Addor et al., 2018; Beck et al., 2015; Kratzert et al., 2019; Kuentz et al., 2017) that found that climate is a stronger control on signatures than landscape. The current manuscript needs to be more precise in stating that "if the scale of the assessed correlations changes, the question changes" (and thus the answer).

We sincerely thank the reviewer for the thoughtful feedback. We especially appreciate the recognition of our alternative approach to LSH, and we fully agree with the need to better highlight how this contrasts with traditional LSH studies. In typical LSH work, correlations are computed across all catchments simultaneously, aiming to identify continental-scale dominant controls. In contrast, our study takes a basin-wise approach, analysing relationships within individual regions/clusters. This shift inherently changes the research question: we are not aiming to assess universal drivers across Europe, but rather to understand what dominates basins in different climates and landscapes.

Additionally, we appreciate the reviewer's suggestion regarding the confounding effect of large climate gradients in LSH. This is a valuable point that supports our argument. We will restructure the introduction stating that climate heterogeneity across large regions may mask landscape effects—further justifying our approach of analysing catchments grouped by region.

- The choice to use Spearman correlations is OK, but not very convincing. I understand that every method comes with its problems (and advantages) but it would be good if the reader gets more confidence that a method that not only considers these individual correlations (but tests for interaction of more catchment attributes) would come to broadly similar conclusions.

We thank the reviewer for raising this important point. We agree that univariate correlation methods have inherent limitations. Particularly in their inability to detect interactions or nonlinear relationships among predictors. Our primary motivation for using Spearman correlations was to maintain interpretability and transparency when comparing across thousands of catchments and multiple geological datasets. Spearman provides a straightforward basis for comparing the strength of individual attribute-signature relationships across scales.

That said, we fully acknowledge that this approach cannot capture more complex interactions between climate, landscape, and geological variables. To address this, we point the reviewer to Section 3.3, where we explicitly discussed this limitation and pointed to previous studies that employed multivariate or machine learning approaches (e.g., Addor et al., 2018; Beck et al., 2015; Kratzert et al., 2019). This helps reassure the reader that while more advanced methods might improve predictive performance, the broad conclusions about the scale-dependent influence of geology and landscape are likely to hold.

- The use of the "maximum |r_s|" seems somewhat odd when not all groups have an equal amount of catchment predictors. Is there a risk that groups that have more predictors are considered to be more dominant, not because they physically are, but because they are more numerous? (I do not expect this to be dominant, but maybe good to consider).

Thank you very much for your comment. There are two key points we would like to clarify:

1. **Group size:** For the geology groups—global, continental, and regional—each contains exactly four variables, corresponding to the four permeability classes. For climate, soils, land use, and topography, the number of variables per group does vary (as shown in Table 3), but the differences are not extreme.

2. **Consequent choice of maximum:** We chose the maximum $|r_s|$ to minimize such differences in group sizes. Our goal was to highlight the highest relationships for each group. Using the median or mean could underrepresent the influence of a group that contains one or two strong, but informative, variables—especially for attributes like geology, where relevant effects may be subtle but specific.

Therefore, we will clarify Section 3.4.1 to acknowledge this trade-off and clarify that while group size may introduce some bias, we judged the maximum $|r_s|$ to be the most informative and interpretable summary metric for this study's purpose.

- Correlation of catchment attributes and signatures seem to be interpreted as "correlation = causation", but that is very speculative. I understand that this often happens in LSH, but here the approach is rather with a scattergun approach: the work studies a very long list of correlations (without physical hypotheses how various factors shape individual catchment attributes) and then picks the strongest correlation at the catchment scale. This seems sensitive to spurious correlations.

Thank you very much for this important comment. It is not our intention to equate correlation with causation, and we fully acknowledge that statistical correlations—especially in large-sample studies—can reflect associations without implying a direct physical mechanism. We also agree that some statements in the original manuscript may have unintentionally given the impression that strong correlations were interpreted as causal relationships. We have carefully reviewed the full text to revise such language and ensure that findings are framed appropriately in terms of statistical association rather than physical inference.

We will also clarify that:

- Spearman correlations are used to highlight potential dominant relationships worth further investigation—not to claim definitive causal links.
- Interpretation of correlations should be made cautiously and considering hydrological expectations and previous literature.

Specific changes will include:

- Rewording sentences in the **Abstract**, **Introduction**, and **Discussion** to avoid causal language (e.g., “control,” “impact”, “causal”) where correlation is meant. Moreover, reword “stronger” to “higher” when discussing correlation values.
- Adding clarifying statements in Section 3.3 and 5.4 that reinforce the distinction between correlation and causation.

Therefore, we believe that with such corrections (see also comments from referee 1), this correlation/causality misinterpretation will be solved, and the manuscript will have a better shape for publication.

- FIg 3: I think it would help if this information was also shown in a different way. For example, scatter plots between geology global values and the other categories. In addition, it would be very helpful it is shown how strongly the best predictors of the different classes are correlated with another (again with scatter plots)

Thank you very much for this constructive suggestion. We agree that comparing geology predictors directly and assessing inter-correlations across attribute groups could be valuable for exploring redundancy or coherence among predictors.

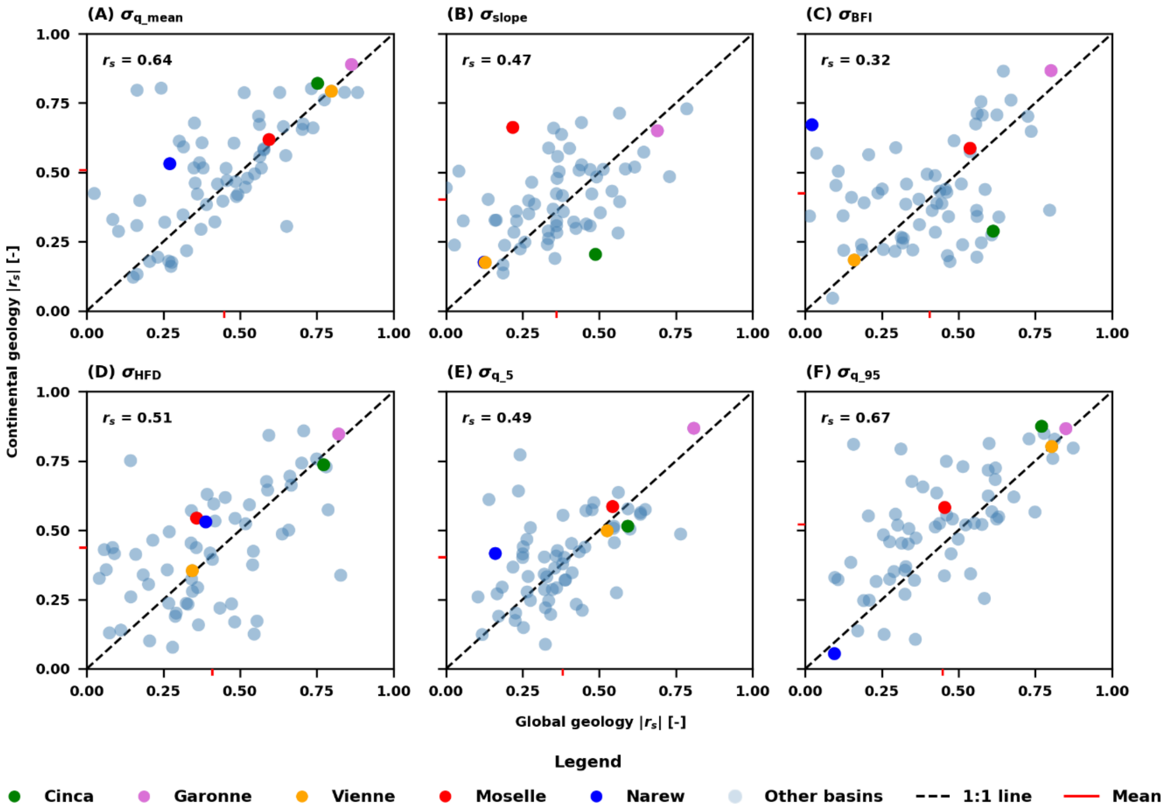
However, after consideration, we believe that including additional scatter plots for all geology predictors (global vs. continental vs. regional) and cross-correlations between the best predictors of each group would introduce significant complexity without proportionate interpretive gain. Since many of the predictors are categorical or derived from reclassifications (e.g., permeability fractions), their relationships are often non-linear, bounded, or collinear by design (summing to 1), which makes such scatter plots potentially misleading or difficult to interpret without heavy qualification.

Moreover, part of the strength of our current approach—especially in Figure 3—is its ability to condense a large volume of correlation results across multiple basins and attribute groups into a format that is both interpretable and comparative. Adding multi-dimensional scatter plots would complicate this clarity and risk distracting from the paper’s core objective: understanding how dominant controls vary with scale and geological map detail.

For these reasons, we have chosen not to add the suggested scatter plots, though we acknowledge the value of such diagnostics and may explore them in a follow-up analysis or supplementary study focused more explicitly on predictor redundancy and interaction.

- Fig. 4: this comparison between the effects (or correlations) of global geology vs continental geology is useful, but it would benifit from shwong a but more than just the data points and 1:1 line. What are the average values of both (you can show the average X and Y coordiante, and this summarizes how well one predicts vs the other, and how well they overall predict). You can also show the correlation (coefficient) between the two data points to show how strongly related they are to another. Such things could greatly help the interpretation of these plots.

  Thank you very much for this comment. These are extremely valuable insights into the figure. We took them into consideration and here you may find an updated version of it.



(A) σ_{q_mean}
r_s = 0.64
(B) σ_{slope}
r_s = 0.47
(C) σ_{BFI}
r_s = 0.32
(D) σ_{HFD}
r_s = 0.51
(E) σ_{q_5}
r_s = 0.49
(F) σ_{q_95}
r_s = 0.67
0.00
0.25
0.50
0.75
1.00
Continental geology $|r_s|$ [-]
Global geology $|r_s|$ [-]
Legend
Cinca
Garonne
Vienne
Moselle
Narew
Other basins
1:1 line
Mean