

We thank the Reviewer for highlighting these crucial points.

Please find our detailed responses below.

Reviewer #1:

This article presents a novel and ambitious analysis of historical surface hydrological connectivity across ten Swiss catchments, using digitized versions of the Siegfried map developed between 1870 and 1926. The authors derive a set of hydromorphological metrics as proxies for surface hydrological connectivity, including drainage density, confluence density, stream order metrics, and shoreline metrics related to different land cover types. These metrics are then statistically related to landscape variables such as lithology, topography, and historical land use/land cover to identify the main drivers of hydrological connectivity in the late 19th century. The study finds that topographic features, particularly valley morphology and slope, are consistent primary drivers of longitudinal connectivity, while wetlands and forested areas play key roles in lateral connectivity. Several unexpected relationships are identified, such as the positive association between building area and hydrological metrics in some catchments, which are interpreted in the context of historical irrigation and land use practices.

I have following comments for improvement before possible publication.

Major Comments

1) The methodology is well developed and generally appropriate for the research questions posed, but there are areas where improvements could increase scientific robustness and interpretive power. The use of the Siegfried map is innovative and provides a good source of historical spatial data, but the temporal spread of the maps across more than four decades introduces a source of temporal inconsistency that is only partially acknowledged. The assumption that these maps provide a coherent "snapshot" of late 19th-century conditions should be supported by more quantitative analysis or stratification. For instance, stratifying map sheets by decade or analyzing whether certain catchments are represented earlier or later in the mapping process could reveal biases in the spatial representation of features. Alternatively, some form of uncertainty band or temporal metadata analysis could be introduced to account for potential inconsistency.

1) Reply

The large temporal spread of four decades was due to the presence of two newer map sheets used to complete the area of the Moesa and Calancasca catchments (see Figure 1 below). Without these two map sheets the total temporal spread total for the 10 catchments together is 26 years (from 1872 to 1898).

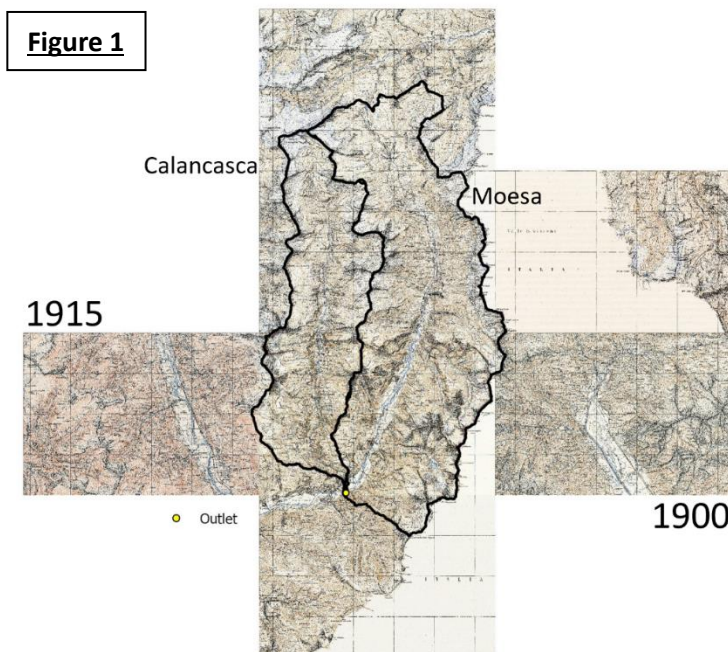
The temporal spread for the catchments is:

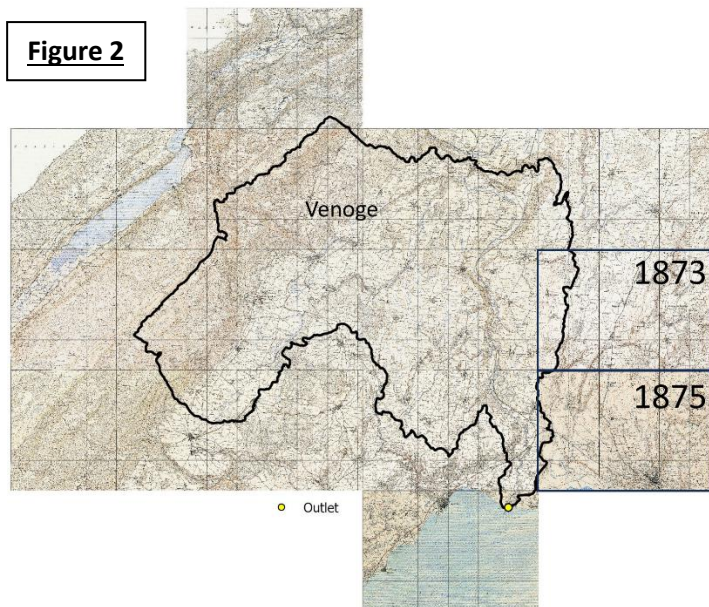
Catchments	Temporal spread	Years
Neker and Glatt	8	1879 - 1886
Aabach and Suhre	13	1878 - 1891
Suze	4	1875 - 1879
Venoge	8 (without 2 marginal older sheets – see Figure 2)	1890 - 1898
Moesa and Calancasca	3 (without 2 marginal newer sheets – see Figure 1)	1872 - 1875
Mattrevispa and Navisence	12	1880 - 1892

We have reformulated the paragraph starting in line 92 in this way:

“To cover the entire catchment area of the ten study catchments, we relied on the oldest available map sheets for each catchment, resulting in a time span of approximately 26 years between the oldest (dated 1872) and the newest map sheet (dated 1898). Please note that two newer map sheets dated 1990 and 1915 were employed to cover a very limited area of the Moesa and Calancasca catchments. While the 26 years’ time span may seem extensive, it is crucial to note that not all map sheets were regularly updated, and updates were carried out in no apparent order (Räth et al., 2023). Consequently, landscape modifications might only appear on updated map sheets years after they occurred. Thus, we interpret the information from these various map sheets as providing an averaged snapshot of the spatial distribution of surface connectivity in the different catchments during the late 19th century. Furthermore, to the best of our knowledge, no major human interventions or extreme natural events (such as landslides and floods) capable of altering the topography of landscapes and streams have affected the ten study catchments during the period considered.”

As noted in line 95, landscape changes may not appear on updated map sheets until years after they occurred. An example is a concrete dam built around 1892 (the "Eberleweiher", in the Glatt catchment), which does not appear on the maps until the 1900 update. Because of these delays, a temporal analysis based on the year of the map sheets would rely on inconsistent timeframes, making it dependent on arbitrary and subjective decisions rather than accurately reflecting when changes actually took place.





2) (I) The derivation of hydromorphological metrics is clear and well justified, yet the redundancy observed between certain metrics, notably the perfect correlation between drainage and shoreline density, suggests a need for dimensionality reduction or orthogonality testing. Methods such as principal component analysis or hierarchical clustering could be used to examine the independence of metrics and reduce redundancy in the dataset. (II) The use of stream order (Strahler) is appropriate, but it would benefit from a sensitivity test given the known limitations of applying stream order logic to historical maps where ephemeral streams might be underrepresented and headwaters potentially truncated due to mapping resolution or classification ambiguity.

2) Reply

(I) Prior to analysis, the non-parametric Spearman's rank correlation test ρ ($p; \alpha = 0.01$) was applied to test the occurrence of monotonic relationships among the metrics and landscape variables. When $\rho = \pm 0.90$ only one variable was retained for further analyses. In the case described by the reviewer, we indeed retained only one variable (i.e., drainage density) for further analysis.

(II) Thank you for raising this important point. We acknowledge the limitations of using Strahler stream order, particularly in historical maps where ephemeral streams may be underrepresented and headwaters potentially truncated due to mapping resolution and classification ambiguities. However, after discussion with our statistician, we came to the conclusion that conducting a sensitivity analysis in this context would require us to make a series of subjective assumptions, such as defining what constitutes an ephemeral stream, estimating their likely locations, and determining the extent of headwater truncation. We believe that these assumptions would introduce significant uncertainty and could bias the results rather than improve their robustness.

Given the constraints of historical map data, we opted to work with the available mapped stream network as consistently as possible, recognizing that any attempt to artificially modify stream order would be speculative.

3) There is limited quantitative treatment of uncertainty in the metric derivation process. While the authors recognize spatial inaccuracies in the historical maps and the potential for scanning and digitization errors, these are discussed only qualitatively. A formal uncertainty propagation analysis—perhaps using Monte Carlo simulations that randomly shift hydrographic features within known error

bounds—would provide confidence intervals for key metrics like drainage density and confluence density. Without this, small observed differences between catchments or sub-catchments may not be statistically meaningful.

3) Reply:

Thank you for your insightful suggestion regarding uncertainty quantification. The extraction of hydrographic features from historical maps in our study is performed using machine learning algorithms, followed by manual correction. We clarified this in the paragraph starting in line 84, which now reads “The following features were extracted from the scanned map through machine-learning algorithms and translated into digital vector data (see Wu et al., 2022 for more details): the hydrographic network (rivers, streams, lakes and wetlands), the forested areas, the area of the glaciers and the buildings’ footprint (i.e. the area on the map covered by the single buildings; Heitzler and Hurni, 2020). All the obtained digital vector data have been subsequently manually corrected to ensure that the extracted vector data had a one-to-one correspondence with what depicted on the map.”

While the initial model introduces some uncertainty, the subsequent manual correction ensures that digitization errors are minimized, effectively eliminating defined error bounds for the extracted features. The primary source of uncertainty, therefore, lies in the original historical maps themselves—their accuracy, resolution, and potential inconsistencies in how hydrographic features were represented at the time of their creation [we discuss this in lines 80 to 83].

While we acknowledge this uncertainty, testing its impact by artificially shifting hydrographic features raises several challenges. First, given that we have the valley bottom delineated, streams could theoretically be placed anywhere within this area, making any adjustment inherently subjective. Second, a Monte Carlo approach would require not only shifting the hydrographic features but also adjusting all spatially dependent landscape variables accordingly. This level of spatial dependency makes it difficult to define a realistic and coherent set of assumptions for the randomization process, potentially introducing more uncertainty rather than clarifying the true variability in our metrics.

4) The classification of landscape features from the historical maps, particularly land use and land cover types such as forests, wetlands, buildings, and fields, appears to rely on machine learning, but little detail is given on the classification model used or its validation accuracy. Accuracy assessment using known control areas or expert interpretation would strengthen the validity of the LULC dataset. This is especially important given that key explanatory variables in the regression models depend heavily on this classification. Misclassification of small but hydrologically important features such as wetlands could lead to erroneous model outputs or misinterpretation of connectivity relationships.

4) Reply:

Similar to our response to question 3, all hydrological and landscape features extracted from the Siegfried map were manually corrected after processing with machine learning algorithms and vectorization. This thorough verification ensures that misclassification is minimized to nearly zero.

5) The multiscale approach is a major strength of the paper, as it acknowledges the hierarchical nature of hydrological systems. The use of mixed-effects models with nested spatial structure and the control of spatial autocorrelation using locational variables is appropriate and well-motivated. (I) However, no diagnostic tests are reported to assess whether spatial autocorrelation remains in model residuals. Simple tests such as Moran’s I or residual correlograms could be used to confirm that spatial dependence has been sufficiently removed, which is crucial for the validity of regression coefficients. (II) In addition, the model structure includes a large number of potential explanatory variables, and while multicollinearity is addressed through VIF analysis, the possibility of overfitting or omitted variable bias remains. Using penalized regression methods or cross-validation could help address this.

5) Reply:

(I) As suggested by the reviewer, we carried out Moran's I test on the residuals of the mixed-effect models. The results show that in some of the models, spatial autocorrelation could still be present in the residuals. However, despite being statistically significant, the obtained Moran's I statistics for the residuals across the models are consistently low (average = 0.00479, standard deviation = 0.00514). This suggests that the models have successfully removed most of the spatial dependency present in the data.

(II) We acknowledge that the model structure includes several potential explanatory variables and, despite we have addressed multicollinearity through VIF analysis, there could be the concern regarding overfitting and omitted variable bias. However, the primary goal of using the regression models in this study is to explore and interpret patterns and trends in studied catchments, rather than to identify a highly optimized model. As such, we deliberately retained as many variables as possible, as all are believed to potentially influence the response variable.

6) The assumption that topography, especially the valley bottom extent and slope, has remained unchanged since the 19th century underlies several parts of the validation and metric alignment. While reasonable in many alpine regions, this assumption may not hold in dynamic sedimentary basins or in areas with known anthropogenic modifications such as channelization or urban expansion. In such cases, the comparison between historical hydrography and contemporary DEM-derived valley bottoms may introduce bias. Some discussion or sensitivity testing around topographic stability over the study period would enhance the reliability of the validation procedure.

6) Reply:

We employed the contemporary DEM to derive the following variables: (sub)catchments' area, mean elevation, mean slope, percentage of area covered by the valley bottom, and mean valley width. Although these variables may be affected by the dynamics in sedimentary basins such as sediment deposition or river incision, in lack of extreme events these processes take a long time to cause noticeable topology changes. Anthropogenic modifications such as channelization or urban expansion can substantially modify the internal structure of the valleys but should not drastically change the general topology of the valley in terms of width and area.

7) The regression framework is sound and includes both fixed and mixed-effect models, with appropriate use of generalized models and two-part modeling for zero-inflated variables. However, the relationships between hydrological metrics and landscape variables are often complex and potentially nonlinear. The current linear framework may not fully capture these complexities. Generalized additive models (GAMs) could be employed to allow for more flexible, non-parametric relationships between predictors and response variables. This may be particularly useful for variables such as stream order or wetland connectivity, which may exhibit threshold or saturation behaviors.

7) Reply:

Thank you for your suggestion regarding the use of GAMs. We recognize that hydrological and landscape relationships can be complex and potentially nonlinear. We also discussed this important point with our statistician. However, applying GAMs effectively typically requires a large and high-quality dataset to robustly estimate smooth functions, which represents a challenge in our case. Our map-derived dataset is inherently noisy, and flexible, data-driven models such as GAMs are prone to overfitting or detecting spurious patterns. Our goal was to employ a model that balances interpretability and complexity—one that is structured enough to account for spatial variables while remaining transparent in its assumptions. By using a simpler regression framework, we aimed to identify broad trends in the data, test our hypotheses, and then validate findings through spatial

analysis on the maps themselves. Introducing more complex models, such as GAMs, could increase the risk of overfitting and drawing inferences that are not strongly supported by the data. Furthermore, even if a GAM revealed a nonlinear relationship, we lack independent, process-based evidence to confirm whether this reflects true ecological mechanisms (e.g., saturation thresholds in wetland connectivity) or is an artifact of data limitations. Without additional field validation, increasing model flexibility may not meaningfully enhance our understanding of the underlying hydrological processes.

We appreciate this suggestion and acknowledge the value of GAMs in certain contexts, but given our dataset's constraints and our study's objectives, we believe our chosen approach provides the most reliable and interpretable insights.

8) While the ecological relevance of hydromorphological metrics is discussed, the connection between these metrics and specific ecological processes or outcomes remains somewhat abstract. The paper could benefit from more explicit hypotheses or examples linking SHC to ecological functions such as habitat continuity, fish migration, or nutrient flux. Even though the historical nature of the data makes direct ecological validation difficult, drawing stronger conceptual links to ecological theory would help bridge the hydrological and ecological dimensions of the study.

8) Reply:

We acknowledge the reviewer's suggestion to establish a more explicit connection between SHC and specific ecological functions. While we agree that such links are important, we believe that a detailed exploration of these ecological processes extends beyond the scope of our manuscript. As the reviewer rightly pointed out, the historical nature of our dataset makes direct ecological validation challenging, and any connections drawn between past SHC conditions and past ecological functions present in the studied catchments would be inherently inferential.

Nevertheless, we have reported the ecological relevance of each SHC metric in Table 2. These descriptions provide insight into how different aspects of connectivity may relate to ecological processes such as habitat provision and complexity, transport of nutrients and organisms, species dispersion and recolonization, and network-scale community stability.

Minor comments:

- **Line 12:** "Human activities have progressively altered..."

We decided to retain the sentence as it is, as we think it better connects to the historical framework of the study.

- **Line 23:** "into large-scale hydrological processes"

We corrected the sentence with "...gain a deeper understanding of important hydrological processes at catchment-scale".

- **Line 41:** instead of 'its vegetation cover' "the vegetation cover."

We corrected the sentence as suggested.

- **Line 62:** "Siegfried map,"

The sentence should be correct without the comma.

- **Line 64:** "primary drivers of surface hydrological connectivity..."

We corrected the sentence as suggested.

- **Table 1 caption (line 70):** "Hypothesized relationships between the landscape drivers..."

The sentence should be corrected without adding "the".

- **Line 88:** "...likely represent the perennial component"

We corrected the sentence as suggested.

- **Line 165:** "Please note that the percentages do not sum up to 100%..." consider placing this in a footnote or table caption to reduce main text clutter.

This information is already part of the table caption and does not appear in the main text.

- Consistency in units and formatting in Table 3 (e.g., km² vs. km2), this applies to other parts in the manuscript.

We uniformed the units as suggested.

- **Line 218:** should be "Figures were generated..."

We corrected the typo.

- **Line 274:** "...to a lesser extent, by the percentage of area covered by buildings..."

We could not understand what the reviewer is suggesting here.

- **Line 286:** "...the average slope mostly along the second axis." → Consider specifying the direction (positive/negative).

To keep consistency within the section, we prefer to not to add this detail here.

- **Table 5:** Ensure consistent use of "Stream order (max)" vs. "Max stream order" and so on across tables and text.

We consistently used stream order (max) in tables and figures and "maximum stream order" in text.