

Maria Grazia Zanoni
Center Agriculture Food Environment- C3A
University of Trento, Trento, Italy
E-Mail : mariagrazia.zanoni@unitn.it

EGU Geoscientific Model Development : Editorial Office

Dear Editor,

We sincerely thank the reviewers for their thoughtful comments and constructive suggestions. We are pleased by their positive feedback on our work and have carefully incorporated their recommendations to improve the clarity and style of the manuscript. With regard to Reviewer 2's request to review the paper again after the revisions, we hope an additional round is unnecessary, as the reviewer provided clear guidance and the required changes were all easily implemented.

We enclose the revised manuscript and provide below a comprehensive, point-by-point summary of all modifications made in response to the reviewers' remarks.

- Reviewer 1 asked for clarifications regarding the specific advantages of using GAMCR over ERRa.

We clarified at the end of the introduction (lines 62–66 in the revised manuscript) that *Differently from ERRa, GAMCR aims to estimate the hydrologic response to each individual precipitation events using combinations of spline basis functions, with coefficients determined through machine learning techniques. This approach, though requiring to fit Generalized Additive Models, allows for greater flexibility since additional information (e.g., temperature, dam operations, or site-specific characteristics) can be incorporated into the model.*

- Reviewer 1 asked for clarifications on the period used for training/testing the model

In Section 3.2 (lines 208 and 209 in the revised manuscript), we clarified that the training/test split for the real data : we used 13 years of data (2005-2017) to train the models. We used data from the years 2018 and 2019 to test the model.

- Reviewer 1 suggested to provide values of hyperparameters λ_1 and λ_2 in section 2.3.

We did the modification in the revised manuscript at line 144.

- The reviewer pointed out that model validation is only introduced later in Section 4, rather than in the methods.

We added a subsection, *Model validation strategy*, to the methods section (Section 3) to clarify our approach from the outset, allowing Section 4 to focus solely on the results.

- Following Reviewer 1's suggestion, we added the link to the GAMCR model software and online tutorial in the text (Section 2.4, line 154 in the revised manuscript).

- The baseline synthetic dataset was generated by fine-tuning the model parameters to approximate the streamflow conditions observed at the Chiasso (Ponte di Polenta) gauging station. Reviewer 1 asked to include quantitative model performance statistics (e.g., R², NSE, PBias, RMSE) for Case A.

We included the RMSE statistic (which is 0.20 mm h⁻¹) in the supplement (line 75) but prefer to omit

it from the main manuscript, as it might mislead readers into thinking the synthetic data generator is designed to best reproduce observed data. Its purpose is instead to create hydrologically realistic series with fully known responses, not to best replicate the specific data at the Chiasso, Ponte di Polenta station.

In Section 3.1 (lines 185 and 186 in the revised manuscript), we also clarify that our goal was to create *a synthetic streamflow time series (...) that realistically reflects observed dynamics (case A) without aiming for exact replication.*

- Reviewer 1 suggested to clarify why selecting a medium-sized catchment was a criterion for the real data.

In Section 3.2 of the revised manuscript (lines 215–219), we clarified this point saying that *we selected these sites because of their diversity in hydrological regimes, elevation and soil depths, which we expect will be reflected in substantially different hydrologic responses. Additional criteria included minimal glacier influence, natural flow regimes (no dams or major abstractions), and complete, reliable data records. The sites have comparable size (between 34–185 km²), which classifies them as small to small/medium basins. Further catchment characteristic analyses appear in the Supplement (Section S1).*

- Following Reviewer 1’s suggestion, we added a Table 2 in Section 3.2 with key hydrological statistics. We also provide the FDCs for each watershed for the snow-free period in a new figure (Figure 5 of the revised manuscript).
- Reviewer 1 pointed out that figure labels a, b and c in Figure 5 (of the original submission) were not matching the Case A, B and C of the synthetic datasets. To avoid confusion, we present Case A, B and C as figures a, b and c respectively in the revised manuscript.
- Reviewer 1 suggested to elaborate on why GAMCR struggles with peak lag prediction. Additionally, we were asked to be more specific about what changes to the basis function might be helpful.

In Section 5 of the revised manuscript (lines 386–392), we added that : *Predicting peak lag is statistically challenging because small changes in lag often cause only minor variations in discharge, making estimation difficult. While larger training datasets can help, model architecture—particularly the choice of basis functions—is crucial. GAMCR’s reliance on unimodal basis functions (namely B-splines with irregular knots) may introduce ambiguity, as dense placement at short lags can bias peak selection due to identifiability issues. Future work could focus on learning basis functions directly from data under suitable constraints, allowing models to adapt flexibly to watershed characteristics while preserving meaningful inductive biases.*

- Following Reviewer 1’s suggestion, we made sure in the revised manuscript to use consistent terminology throughout the paper. More precisely, we stick to the terminology *training* and *test periods* to enhance clarity.
- Reviewer 2’s primary concern was that the method was presented as globally applicable despite being calibrated and validated only in Swiss watersheds. Reviewer 2 recommended either limiting the claims of applicability or extending validation to more diverse basins worldwide.

We revised the manuscript to restrict claims to basins with similar characteristics to those tested here, and we clarified that broader validation in other regions is acknowledged as an important direction for future research (lines 403–406 in Section 5 in the revised manuscript). We also added new analyses (Flow Duration Curves and runoff coefficients) to highlight the diversity of the tested sites.

In our response to the reviewer, we acknowledged that, in a global context, the selected Swiss basins are not *climatically* diverse. However, within Switzerland, the basins were chosen to cover a range of hydrological

regimes, including Jura-nivopluvial, transition nival, pluvial, southern nivo-pluvial, and southern pluvio-nival. This diversity allowed us to test the GAMCR method across different runoff behaviors and seasonal patterns resulting in a broad range of responses. Indeed, Figure 9 of the revised manuscript shows that the same rainfall intensities result in hydrologic responses that are ten times larger at Euthal compared to Salmsach.

We believe that our site selection does not diminish the value of the results, as the methodology and framework developed with these basins are applicable to other regions, provided that careful data evaluation is conducted. Testing in other climates and continents is an important avenue for future work, but it lies beyond the scope of this initial study.

- Following Reviewer 2’s suggestion, we clarified the abbreviation GAMCR in the abstract of the revised manuscript (lines 7 and 8).
- Following Reviewer 2’s suggestion, we inserted references for the sentences *the hydrologic response (or runoff response) is usually defined as the change in streamflow induced by a given input of precipitation.* (line 20 in the revised manuscript) and *In the case of the hydrologic response, the evaluation step is particularly important because the impulse response functions cannot be measured directly* (line 174 in the revised manuscript).

- Reviewer 2 asked us to clarify the statement suggesting that the method "varies smoothly between basins with different climatic and physiographic characteristics."

We answered that GAMCR is trained and applied within a specific catchment, and the intended meaning was that hydrologic responses vary over time in a continuous, not abrupt, manner. We rephrased the sentence (in Section 2.2, lines 107–108) to : *This means we expect similar feature vectors to produce similar hydrologic responses.*

- Reviewer 2 argued that basins of $34 - 195 \text{ km}^2$ should be classified as small to small/medium rather than medium-sized, and stressed that this highlights the need to test the method in other basin types.

We answered that we agree with the area-based classification but note that there is no universally accepted standard, as some hydrologic definitions emphasize process-based criteria rather than size alone. Since the studied basins meet such process-based characteristics of midsize catchments, we revised the manuscript (Section 3.2, lines 215–219) to clarify both perspectives-area-based and process-based to avoid ambiguity and better justify our methodological choices.

- We fixed some typos gently notified by the two Reviewers.
- The reviewers raised a few additional points that did not require changes to the manuscript, but we have addressed them in the discussion.
 - Reviewer 1 shed light on the modifications to the Swiss radar network implemented between 2005 and 2016, which significantly improved the data quality of the CombiPrecip product. We explained that we included data starting from 2005, as initial experiments indicated that the model’s performance was not substantially impacted by the lower data quality prior to 2015. This decision also reflects a more realistic scenario in which flux data is available over at least a decade.
 - Reviewer 1 highlighted that the NSE for the out-of-sample prediction for the Chiasso dataset is 0.19, and asked to clarify why the model shows such low NSE for this real dataset, despite the apparent

agreement between the synthetic and observed training data. We explained that the NSE coefficient is known to be sensitive to high-flow values, and a few large errors-typically occurring under high-streamflow conditions, and possibly affected by poor rainfall estimates-can markedly reduce NSE values, even when the model performs well in most other cases.

- Reviewer 2 noted that antecedent conditions in the studied basins are influenced by snow, even outside the snowy months, and emphasized this as another reason to test the method in other watershed types.

We answered that we recognize this influence and that we had indeed selected events and basins carefully to minimize snow-related effects, ensuring a robust evaluation under the targeted conditions. While this limits the generality of the results, it avoids confounding factors such as snowmelt. We agree that extending validation to basins with different snow or climatic regimes is important future work, though beyond the scope of this study.

Yours sincerely,

M.G. Zanoni on behalf of the co-authors