nature portfolio

Peer Review File

Hybrid approaches enhance hydrological model usability for local streamflow prediction

Corresponding Author: Professor Ilias Pechlivanidis

This file contains all editorial decision letters in order by version, followed by all author rebuttals in order by version.

Version 0:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Professor Pechlivanidis,

Your manuscript titled "Hybrid approaches enhance hydrological model usability for local streamflow prediction" has now been seen by 2 reviewers, and we include their comments at the end of this message. They find your work of interest, but some important points are raised. We are interested in the possibility of publishing your study in Communications Earth & Environment, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response that takes into account the points raised. Please highlight all changes in the manuscript text file.

Please submit your point-by-point responses as a separate file, distinct from your cover letter where you can add responses to the Editors' comments that you do not want to be made available to the reviewers. Word files are preferred. We recommend that any figures, tables or graphs that are included in the response to reviewers are also included in the main article or Supplementary Information.

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the referees' comments (which should be in a separate document to any cover letter), a tracked-changes version of the manuscript (as a PDF file) and the completed checklist:
Link Redacted
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to receive your revised paper within six weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

Please do not hesitate to contact us if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Somaparna Ghosh, PhD
Associate Editor,
Communications Earth & Environment

EDITORIAL POLICIES AND FORMATTING

We ask that you ensure your manuscript complies with our editorial policies. Please ensure that the following formatting requirements are met, and any checklist relevant to your research is completed and uploaded as a Related Manuscript file type with the revised article.

Editorial Policy: Policy requirements (Download the link to your computer as a PDF.)

For Manuscripts that fall into the following fields:
• Behavioural and social science
• Ecological, evolutionary & environmental sciences
• Life sciences
An updated and completed version of our Reporting Summary must be uploaded with the revised manuscript
You can download the form here:
https://www.nature.com/documents/nr-reporting-summary.zip

Furthermore, please align your manuscript with our format requirements, which are summarized on the following checklist:
Communications Earth & Environment formatting checklist

and also in our style and formatting guide Communications Earth & Environment formatting guide .

*** DATA: Communications Earth & Environment endorses the principles of the Enabling FAIR data project (http://www.copdess.org/enabling-fair-data-project/). We ask authors to make the data that support their conclusions available in permanent, publically accessible data repositories. (Please contact the editor if you are unable to make your data available).

All Communications Earth & Environment manuscripts must include a section titled "Data Availability" at the end of the Methods section or main text (if no Methods). More information on this policy, is available at http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf.

In particular, the Data availability statement should include:
- Unique identifiers (such as DOIs and hyperlinks for datasets in public repositories)
- Accession codes where appropriate
- If applicable, a statement regarding data available with restrictions
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Data Availability Statement.

DATA SOURCES: All new data associated with the paper should be placed in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific, community-recognized repositories, where possible and a list of recommended repositories is provided at http://www.nature.com/sdata/policies/repositories.

If a community resource is unavailable, data can be submitted to generalist repositories such as figshare or Dryad Digital Repository. Please provide a unique identifier for the data (for example a DOI or a permanent URL) in the data availability statement, if possible. If the repository does not provide identifiers, we encourage authors to supply the search terms that will return the data. For data that have been obtained from publically available sources, please provide a URL and the specific data product name in the data availability statement. Data with a DOI should be further cited in the methods reference section.

Please refer to our data policies at http://www.nature.com/authors/policies/availability.html.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

The paper “Hybrid approaches enhance hydrological model usability for local streamflow prediction” has been submitted to the journal “communications earth & environment” for peer review. The paper claims to enhance the quality of streamflow simulations from large-scale hydrological models (LSHMs) in a way that allows for more reliable applications at the local scale. They claim to do this by answering three questions:

- How can hybrid process-based and statistical/machine-learning models be effectively integrated to enhance local-scale hydrological predictions across diverse hydrological regimes?
- What mechanisms underpin the spatial complementarity of post-processing methods?

- How can this inform tailored strategies for optimizing hydrological model performance in different regions?

The authors also claim that their research can help with predictions in ungauged basins and for flow forecasting applications.

The submission explores four different post-processing techniques (2 statistical and 2 machine-learning) to improve streamflow predictive capacity. The authors claim that their work enhances streamflow prediction with LHSMs with socio-economic implications to multiple sectors such as agriculture, hydropower, and drinking water. The paper then presents the authors' rationale for answering the three questions they pose.

I believe that this work is novel and would be of interest to others in the community, particularly as more and more researchers blend physically based modelling with statistical and machine learning approaches.

However, there are significant improvements that can be made to the manuscript, as described in my general comments below. None of these improvements should necessitate any additional model runs but would benefit from adding discussion. Therefore, I recommend to the journal editors that the authors revise their manuscript to address these comments before a final decision is reached. I would be happy to review a revised manuscript.

My general comments are as follows:

Major Comment

My major concern with the paper is that the authors don't provide convincing arguments that their work addresses their questions and claims. It is not clear to me that they have answered their own questions, nor their claims that their methods can be used in ungauged basins or for flow forecasting applications. This concern can likely be addressed by either reframing their questions and claims, or by providing further discussion around how the research can answer these questions and support their claims.

Overall Organization of the Paper

The overall organization of the paper should be improved for readability. Although not required at this stage, I recommend that the authors follow the journal's style and formatting guide found at https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf. In particular, the following alterations would improve the readability:

- A paragraph should be added at the end of the introduction (main) to provide a summary of the major results and conclusions.
- A slightly longer summary of the work completed (methods) should be provided in the introduction (or the "main" section as submitted), including a Figure showing the domain (Europe) and the stations used in the evaluation. Extended Data Figure 7a could be used for this purpose and removed from Extended Data Figure 7, which would more effectively set the stage for the reader at the outset.
- I recommend that the results and discussion be split into two sub-headings (results, then discussion) for the sections associated with each of the three questions. Most of what is presented in the current manuscript could be the results with more discussion included for each set of results.
- I find the methods section in need of additional detail, especially if the reader is interested in reproducing the results or performing a similar analysis on different datasets. More detailed feedback is provided at the end of this review.
- Please include a "Data availability" and "Code availability" sections as required by the style and formatting guide. In my experience, the earlier that an author thinks through data availability, code availability and reproducibility issues, the more likely others will be able to reproduce the results or perform similar analysis. The ideal would be for the scripts and code used to download and analyze the underlying datasets to be available on a GitHub repository or something similar.
- All figures should be described in their captions more effectively so that the reader can more easily understand what each figure is showing. Many Figures also contain text in a very small font that is difficult to read even when magnified on a large computer screen. This should be fixed.
- It is not clear where in the manuscript the authors answer the three questions they pose in the introduction (main). Based on the titles of the subsequent sections, I assume that each subsequent section should answer one of the questions. But they don't. The questions are also not answered in the sub-section "Enabling knowledge transfer for a wider global impact" prior to the methods section.

Section 1: Hybrid modelling improves performance at local scale.

- The first section "Hybrid modelling improves performance at local scale" doesn't answer the 1st question "How can hybrid process-based and statistical/machine-learning models be effectively integrated to enhance local-scale hydrological predictions across diverse hydrological regimes?" This section of the manuscript simply shows that the 4 post-processing methods used in the study improve the results. This should not be surprising at all, considering that improving the results is the reason one wants to post-process the results in the first place.
- Most of this section shows the results and explains what the results show, without discussing the details about how the results are unique and why they are important. This section could be improved by further discussing how the authors have done something novel and advanced the science in a way that hasn't been done before. The meaning of the results should also be further discussed.
- Either the question needs to be reframed, or the authors need to indicate how their work answers their first question.

- I recommend that the discussion for this section conclude with a precise answer for how the authors answer the first question.
- Figure 2 is a sub-figure of Extended Figure 1. Figure 2 can therefore be removed and the text re-written to focus on Extended Figure 1. One of the Chord Diagrams (a sub-figure in Figure 2 or Extended Figure 1) should be more explicitly explained in the methods section.
- The terms “very good, good, fair, poor, very poor, and unsatisfactory” shown in Figure 2 and Extended Figure 1 appear to be arbitrary and are not explained in the methods section. What was the criteria used to determine what score corresponds to which term? I would refer the authors to Moriasi et al. (2015) for a published example and Clark et al. (2021) for a critical evaluation of the use of popular performance metrics in hydrologic modeling.

Section 2: no single best hybrid method: spatial complementarity of post-processing potential.
- The second section does not answer the question “What mechanisms underpin the spatial complementarity of post-processing methods?” The text describes where each of the methods perform the best without saying very much about what underlying mechanisms explain the results. It’s not clear what “mechanisms” the authors are referring to in the question. Are the mechanisms the physical hydrological processes that are dominant in different cases where one post-processing method performs better than the others?
- Most of this section shows the results and explains what the results show, without discussing the details about how the results are unique and why they are important. This section could be improved by further discussing how the authors have done something novel and advanced the science in a way that hasn’t been done before. The meaning of the results should also be further discussed.
- Either the question needs to be reframed, or the authors need to indicate how their work answers their second question.
- Figure 3 is a sub-component of Extended Figure 2. Figure 3 can therefore be removed and the text re-written to focus on Extended Figure 2.
- Extended Data Figure 3 is the first mention of hydrological clusters. I recognize that these clusters are defined in the methods section but should be briefly noted and cited in the results section. Or perhaps summarized and cited in the introduction so that the reader has a sense of what a cluster is when first coming across its use in the text.
- Is the “local” portion of the manuscript’s title accurate given the statement on lines 187 to 190: “in few river systems, no hybrid method adds value. These stations … being characterized by small upstream area (mostly less than 500 km2)…” The word “local” is used 28 times in the manuscript but is not defined. What exactly is “the local scale” and “local decision making”? The term “local” needs to be precisely defined in the context of the paper. If small upstream areas less than 500 km2 are not local, then what is local?

Section 3: Hydrological regime as a key driver to model performance enhancement.

- Section 3 does not answer the question “How can this inform tailored strategies for optimizing hydrological model performance in different regions?” There is very little discussion around this question with the bulk of the text simply explaining the results.
- The methods section would benefit from more clearly explaining how to interpret the figures in this section.

Section 4: Enabling knowledge transfer for a wider global impact.

- The authors claim in this section that their research is applicable to ungauged locations (lines 280 to 294), but this is not clear at all. Further discussion around this claim is necessary to convince the reader of the claim.
- The authors also appear to claim that their research can support the forecasting of floods and droughts (lines 296 to 308). Again, this is a broad statement that contains very little discussion as to how flow forecasting agencies might be able to use the research. Further discussion is required.

Methodology

The methodology section is lacking key details.
o A detailed description of all diagrams shown in the analysis would be helpful.
o More details around the post-processing methods would also be helpful.
o Regarding the scripts and datasets that were used to perform the analysis. Are the scripts available on GitHub and the data easy to access? This would support reproducibility of the results or expansion of the science to others interested in trying the approaches elsewhere.
o The Figures are not clearly explained and there is some duplication that can be removed.

Clark, M., P., Vogel, R., M., Lamontagne, J., R. et al. (2021). The abuse of popular performance metrics in hydrologic modelling. Water Resource Research. 57, e2020WR029001. https://doi.org/10.1029/2020WR029001

Moriasi, D., N., Gitau, M., W., Pai, N., Daggupati, P. (2015). Hydrologic and water quality models: performance measures and evaluation criteria. Transactions of the ASABE. 58(6): 1763-1785. https://doi.org/10.13031/trans.58.10715

Reviewer #2 (Remarks to the Author):

Review report for ´Hybrid approaches enhance hydrological model usability for local streamflow prediction’ by Yiheng Du and Ilias G. Pechlivanidis

This manuscript demonstrates how hybrid approaches can improve hydrological simulations through incorporating additional post-processing steps based on machine learning techniques such as random forests, long short-term memory models but more straightforward techniques such as quantile mapping and generalised linear model. While the promising effect of machine learning and hybrid approaches have been shown throughout the field in various studies with different angels and interpretation of hybrid version, this study highlights the potential of including these methods for post-processing at the European scale, which has not been done before. The post-processing approaches were combined with the E-HYPE model, a process-based large-scale hydrological model over the entire European domain. The manuscript, highlights not only the added benefit that the tested methods bring for total volume, high and low flow extremes, but also identifies the potential drivers of the performances, which gives an additional interesting insight to the reader. Overall, this manuscript and it's extensive comparison of over 2000 gauging stations throughout whole Europe and the different post-processing approaches creates an interesting overview of this type of hybrid modelling at a large scale.

While the manuscript is well written, a few comments/suggestions/questions arose:
- Consider adding post-processing in the title to make it more clear to the reader what to expect in terms of hybrid modelling
- For readability, consider moving the extended data figures (e.g. Extended Data Fig 3, etc.) to the end or appendix if this would not be already the case in the final edit. While these figures include additional information, the current amount of figures is disrupting the reading flow and distracts from the main highlights discussed in the different sections (e.g. Fig 2, Fig 3 and Fig 4b feature importance)
- Figure 2: it would be interesting to know which stations in which clusters are improving. Or if it's possible to link it to the clusters? Is there a trend which cluster profits most from the post-processing?
- Extended Data Fig 1, not better to link it to Figure 2 as Extended Data Figure 2?
- Figure 3 nice overview with different post-processing methods
- Extended Data Figure 3b) can be confusing if it comes before introducing the clusters. In general, the figure descriptions in the text or captions could be extended in some cases
- Line 232: consider introducing the figure with the clusters from the material and methods also here as the clusters seem to play an important role in the results. Would help the reader follow along better
- Figure 4, Extended Data Figure 4 and Figure 5: consider highlighting the feature importance more as these can be easily interpreted on its own or extend the explanation on the interpretation of the skill drivers for easier interpretation
- Extended Data Figure 5: consider adding total volume, high flow extreme and low flow extreme to the titles of the evaluation scores.
- Methods: Extended Data Figure 6 would be nice to have in the main body to highlight the flow of the work as it gives a nice overview
- Methods: post-processing methods, strongly suggest to highlight that currently it seems that a model per location was developed for the post-processing (not really mentioned otherwise). Furthermore it would be great if the input data for the post-processing step could be clearly stated (currently assuming it's equal to the ones used for the driver analysis?) Line 391: are the listed steps for training, testing and validation all the same for all the methods?
- While in the manuscript it was mentioned that the insights can be useful for ungauged areas as well, I was wondering whether the authors also potentially considered making one (e.g. LSTM) per cluster or the whole region? In other studies from the field, there's a discussion that machine learning models trained on various catchments do better for ungauged catchments, which could also be interesting here if the post-processing would be planned to be extended for other locations as well with less observational data for model training.
- Table 3: slope replaced by precipitation?
• What are the major claims of the paper?
• Are the claims novel? If not, please identify the major papers that compromise novelty.
• Will the paper be of interest to others in the field?
• Will the paper influence thinking in the field?
• Are the claims convincing? If not, what further evidence is needed?
• Are there other experiments that would strengthen the paper further? How much would they improve it, and how difficult are they likely to be?
• Are the claims appropriately discussed in the context of previous literature?
• If the manuscript is unacceptable in its present form, does the study seem sufficiently promising that the authors should be encouraged to consider a resubmission in the future
• Is the manuscript clearly written? If not, how could it be made more accessible?
• Could the manuscript be shortened to aid communication of the most important findings?
• Have the authors done themselves justice without overselling their claims?
• Have they been fair in their treatment of previous literature?
• Have they provided sufficient methodological detail that the experiments could be reproduced?
• Is the statistical analysis of the data sound?
• Should the authors be asked to provide further data or methodological information to help others replicate their work? (Such data might include source code for modelling studies, detailed protocols or mathematical derivations).
• Are there any special ethical concerns arising from the use of animals or human subjects?

** Visit Nature Research's author and referees' website at www.nature.com/authors for information about policies, services and author benefits**

Communications Earth & Environment is committed to improving transparency in authorship. As part of our efforts in this direction, we are now requesting that all authors identified as ‘corresponding author’ create and link their Open Researcher and Contributor Identifier (ORCID) with their account on the Manuscript Tracking System prior to acceptance. ORCID helps the scientific community achieve unambiguous attribution of all scholarly contributions. You can create and link your ORCID from the home page of the Manuscript Tracking System by clicking on ‘Modify my Springer Nature account’ and following the instructions in the link below. Please also inform all co-authors that they can add their ORCIDs to their accounts and that they must do so prior to acceptance.
https://www.springernature.com/gp/researchers/orcid/orcid-for-nature-research

For more information please visit http://www.springernature.com/orcid

If you experience problems in linking your ORCID, please contact the Platform Support Helpdesk.

Version 1:

Decision Letter:

** Please ensure you delete the link to your author home page in this e-mail if you wish to forward it to your coauthors **

Dear Professor Pechlivanidis,

Your manuscript titled "Hybrid approaches enhance hydrological model usability for local streamflow prediction" has now been seen by our reviewers, whose comments appear below. In light of their advice we are delighted to say that we are happy, in principle, to publish a suitably revised version in Communications Earth & Environment. provided you fully describe how hybrid modelling approach can be used to improve forecasts or to improve predictions in ungauged basins, as suggested by Reviewer #1.

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table".

*****Please take care to match our formatting and policy requirements. We will check revised manuscript and return manuscripts that do not comply. Such requests will lead to delays. *****

Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files as a Related Manuscript file.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at https://www.nature.com/documents/commsj-file-checklist.pdf .

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication. For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit https://www.nature.com/commsenv/open-access

At acceptance, you will be provided with instructions for completing the open access licence agreement on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:

Link Redacted
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Somaparna Ghosh, PhD
Associate Editor,
Communications Earth & Environment
Consulting Editor,
Communications Sustainability

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

I would like to thank the authors for their revision of their paper "Hybrid approaches enhance hydrological model usability for local streamflow prediction." The authors have effectively responded to the majority of my comments as well as the comments of the other reviewer. The paper has, in my view, been improved considerably as a result.

However, it is still unclear to me how the hybrid modelling approach can be used to improve forecasts or to improve predictions in ungauged basins.

Perhaps my lack of clarity stems from not being deeply involved in the hybridization of hydrologic models with machine learning, but I believe that the manuscript could more clearly articulate how such a hybrid approach could be operationalized for someone with my background, which is more focused on hydrologic modelling.

For example, with respect to improving hydrological forecasts, the added text starting on line 408 of the track-change version of the manuscript says that biases between model simulations and observations "can be mitigated by incorporating our proposed modelling framework..." I can understand how this is true for a hind-cast, but not a forecast. How can you adjust errors in a forecast when you don't have observed streamflow (because the future is yet to come)? Are the model parameters somehow changed by the post-processing? Or is the forecasted streamflow somehow changed by the post-processing being applied to the forecasted streamflow without observed streamflow? A few more details would be appreciated.

With respect to ungauged basins, a little more detail is provided, but perhaps not quite enough for someone who isn't doing this exact kind of work on a regular basis. Based on the authors' responses to the other reviewer, it appears that they are working on a regionalization approach for ungauged basins, so perhaps this is worth mentioning in the manuscript along with a few more details.

If the audience of the paper is limited to those who are already heavily involved in the hybridization of hydrologic models with machine learning, then perhaps not much more explanation is needed. For a broader audience, however, clear explanations of how the approach can be operationalized for ungauged basins or forecasting purposes should be included.

If the journal editors wish, I would be happy to review any additional changes made by the authors as this is a very interesting topic of real potential value for operational forecasting agencies around the world.

Reviewer #2 (Remarks to the Author):

The authors have incorporated the feedback of both reviewers diligently, which is much appreciated, and the revised manuscript with the additions and changes made to figures, sections and clarifications improved the flow and the depth of the manuscript. There are no additional remarks.

RESPONSE LETTER

We sincerely thank all the reviewers for the constructive comments. The revision process has been a great learning experience, and it's a pleasure for us to communicate the improvements we have made. Below, we provide detailed responses to each point, highlighted in blue. All line numbers refer to the **track-change version of the manuscript**, with revised text shown in *italics and underlined*.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

The paper "Hybrid approaches enhance hydrological model usability for local streamflow prediction" has been submitted to the journal "communications earth & environment" for peer review. The paper claims to enhance the quality of streamflow simulations from large-scale hydrological models (LSHMs) in a way that allows for more reliable applications at the local scale. They claim to do this by answering three questions:

- How can hybrid process-based and statistical/machine-learning models be effectively integrated to enhance local-scale hydrological predictions across diverse hydrological regimes?
- What mechanisms underpin the spatial complementarity of post-processing methods?
- How can this inform tailored strategies for optimizing hydrological model performance in different regions?

The authors also claim that their research can help with predictions in ungauged basins and for flow forecasting applications.

The submission explores four different post-processing techniques (2 statistical and 2 machine-learning) to improve streamflow predictive capacity. The authors claim that their work enhances streamflow prediction with LHSMs with socio-economic implications to multiple sectors such as agriculture, hydropower, and drinking water. The paper then presents the authors' rationale for answering the three questions they pose.

I believe that this work is novel and would be of interest to others in the community, particularly as more and more researchers blend physically based modelling with statistical and machine learning approaches.
However, there are significant improvements that can be made to the manuscript, as described in my general comments below. None of these improvements should necessitate any additional model runs but would benefit from adding discussion. Therefore, I recommend to the journal editors that the authors revise their manuscript to address these comments before a final decision is reached. I would be happy to review a revised manuscript.

Authors Response:
We appreciate the reviewer's recognition of the novelty and relevance of our work. The willingness to evaluate our revised manuscript is also encouraging. We have addressed all the comments and provide detailed responses below. We are grateful for the reviewer's time and effort in helping us improve the quality of this manuscript, and look forward to further discussions on the revised version.

My general comments are as follows:
Major Comment
My major concern with the paper is that the authors don't provide convincing arguments that their work addresses their questions and claims. It is not clear to me that they have answered their own questions, nor their claims that their methods can be used in ungauged basins or for flow forecasting applications. This concern can likely be addressed by either reframing their questions and claims, or by providing further discussion around how the research can answer these questions and support their claims.

Authors Response:
We agree that strengthening the connection between the scientific questions and their corresponding answers will greatly improve the manuscript. This comment helps us realize that the original scientific questions were somewhat vague. We have refined them to align more precisely with each key result.

- How do hybrid process-based and statistical/ML methods enhance local model performance across various streamflow characteristics?

Answering by the section “Hybrid modelling improves representation of streamflow characteristics at local scale”.

- How does the performance of different post-processing methods vary across Europe’s hydro-climatic gradient?

  Answering by the section “No single best hybrid method: spatial complementarity of post-processing potential”.

- What are the key drivers controlling the hybrid model performance enhancement across Europe?

  Answering by the section “Hydrological regime as a key driver to model performance enhancement”.

Actions in the paper:
Line 81, “*Here, we enhance the quality of streamflow simulations from LSHM at the local scale across the pan-European domain, and improve the understanding of model enhancement to allow for more reliable applications for local decision-making, by answering the following scientific questions: (1) How do hybrid process-based and statistical/ML methods enhance local model performance across various streamflow characteristics? (2) How does the performance of different post-processing methods vary across Europe’s hydro-climatic gradient? and (3) What are the key drivers controlling the hybrid model performance enhancement across Europe?*”

Overall Organization of the Paper
The overall organization of the paper should be improved for readability. Although not required at this stage, I recommend that the authors follow the journal’s style and formatting guide found at https://www.nature.com/documents/commsj-phys-style-formatting-guide-accept.pdf.
Authors Response:
Thanks for the instructions. We have restructured the manuscript to follow the formatting guidelines, organizing it into four sections: Introduction, Results, Discussion, and Methodology. In the original version, the Results and Discussion sections were not stated or separated, which made it challenging to connect the results to the scientific questions as the reviewer pointed out earlier in the previous comment. This has now been addressed.

In particular, the following alterations would improve the readability:

- A paragraph should be added at the end of the introduction (main) to provide a summary of the major results and conclusions.

Authors Response:

We have addressed this point as suggested by adding a concluding paragraph.

Actions in the paper:

Line 98: "*Our analysis, covering over 2000 gauging stations across a wide range of hydrological regimes (Fig. 1b), shows that the two ML methods yield higher improvements than the statistical methods, particularly in capturing extreme streamflow characteristics. However, no single method consistently outperforms across the entire domain, rather a spatial complementarity occurs, which is primarily influenced by catchment characteristics, including hydrological regimes (represented by predefined hydrological clusters[30,32], details in Methodology, Fig. S1 and Table S1 in the supplementary information) and climate conditions among the investigated potential drivers.*"

- A slightly longer summary of the work completed (methods) should be provided in the introduction (or the "main" section as submitted), including a Figure showing the domain (Europe) and the stations used in the evaluation. Extended Data Figure 7a could be used for this purpose and removed from Extended Data Figure 7, which would more effectively set the stage for the reader at the outset.

Authors Response:

We have addressed this comment and integrated it with the revisions made in response to the previous comment. A paragraph has been added to the Introduction, outlining the main workflow of the study, the methods used, and the key conclusions. Following the suggestion, the figure containing the workflow and station locations has been moved from the Methodology to the main text.

Actions in the manuscript:

Line 92: "*To address these questions, we establish a hybrid framework (Fig. 1a) for post-processing the outputs from the E-HYPE process-based LSHM across the entire European domain. This framework employs two statistical methods (Generalised Linear Model, GLM; Quantile Mapping, QM; Methodology section) and two ML-based methods (Random Forest, RF; Long Short-Term Memory model, LSTM; Methodology section), with comprehensive evaluation metrics and performance attribution, allowing a thorough assessment and process understanding. Our analysis, covering over 2000 gauging* *stations across a wide range of hydrological regimes (Fig. 1b), shows that the two ML methods yield higher improvements than the statistical methods, particularly in capturing extreme streamflow characteristics. However, no single method consistently outperforms across the entire domain, rather a spatial complementarity occurs, which is primarily influenced by catchment characteristics, including hydrological regimes (represented by predefined hydrological clusters [30,32], details in Methodology, Fig. S1 and Table S1 in the supplementary information) and climate conditions among the investigated potential drivers. From an operational perspective linked to either early warning systems or climate services, this effort strongly enhances LSHM usability for streamflow predictions at local conditions and carry significant implications for water-dependent sectors (e.g. agriculture, hydropower, drinking water etc.).*"

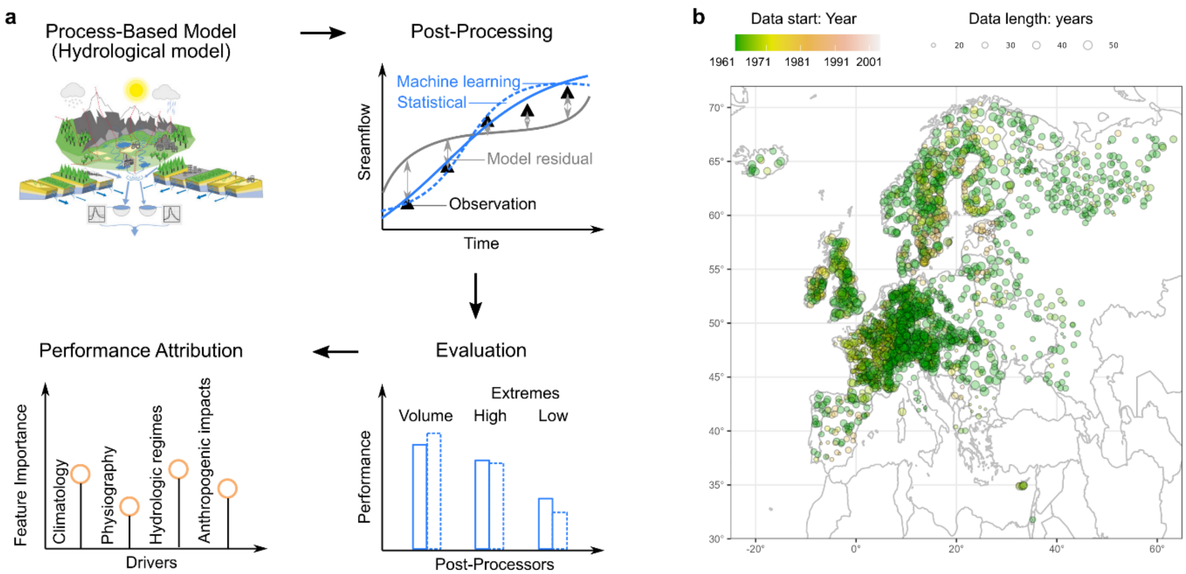


Fig. 1: Hybrid framework for post-processing process-based LSHM and data availability of the observations. *a, Schematic of the hybrid framework, detailing the process-based model, post-processing, evaluation, and performance attribution steps. b, Spatial distribution of the stations used in the study, annotated with the start year and duration of the observational data.*

- I recommend that the results and discussion be split into two sub-headings (results, then discussion) for the sections associated with each of the three questions. Most of what is presented in the current manuscript could be the results with more discussion included for each set of results.

Authors Response:

We agree with the reviewer and have revised the structure accordingly. We also revised the scientific questions to be closely associated with each key message in the Results section. In the Discussion, we have elaborated on two key aspects: improving model accuracy (addressing the first and second scientific questions) and understanding the potential drivers (related to the third question). This ensures consistency throughout the manuscript, also strengthening the connections across sections.

- I find the methods section in need of additional detail, especially if the reader is interested in reproducing the results or performing a similar analysis on different datasets. More detailed feedback is provided at the end of this review.

Authors Response:

Thank you for this suggestion. We have revised the Methodology section by including more detailed explanations of the methodologies used, restructuring certain parts for better clarity, and adding information on the visualizations. Moreover, we have provided codes for the modelling and plotting, making them openly available for download.

Actions in the paper:

Line 462 to 651 in Methodology section.

- Please include a "Data availability" and "Code availability" sections as required by the style and formatting guide. In my experience, the earlier that an author thinks through data availability, code availability and reproducibility issues, the more likely others will be able to reproduce the results or perform similar analysis. The ideal would be for the scripts and code used to download and analyze the underlying datasets to be available on a GitHub repository or something similar.

Authors Response:

Yes, this has been implemented. We have provided a link to Zoneto, where the code for post-processing methods and visualisation, and corresponding data are available for download and use.

Actions in the paper:

Line 676: "*Data availability: We performed the investigations at the SMHI Hydrology Research unit, where work benefits from joint efforts in developing models and concepts by the whole team. The HYPE model code is available from the HYPEweb portal (https://hypeweb.smhi.se/model-water/). Streamflow data, including E-HYPE model simulations and observations, can be shared upon request, following the SMHI data-*

sharing policy. Evaluation metrics and model skill assessments can be accessed in the repository on Zenodo [https://zenodo.org/records/14938526].
Code availability: All plots and post-processing models were generated using the R programming language. The code, including scripts for data processing, model training and evaluation, and visualization, can be found in the Zenodo repository [https://zenodo.org/records/14938526]."

- All figures should be described in their captions more effectively so that the reader can more easily understand what each figure is showing. Many Figures also contain text in a very small font that is difficult to read even when magnified on a large computer screen. This should be fixed.

Authors Response:

Thank you for the comment. As suggested, we have carefully checked all the figures, improved the description of the captions, and increased the text to proper size.

- It is not clear where in the manuscript the authors answer the three questions they pose in the introduction (main). Based on the titles of the subsequent sections, I assume that each subsequent section should answer one of the questions. But they don't. The questions are also not answered in the sub-section "Enabling knowledge transfer for a wider global impact" prior to the methods section.

Authors Response:

Thanks for highlighting this. As noted in our response to the major comment, we consider this an important point and have addressed it accordingly. We recognized the lack of clear correspondence and have refined the scientific questions to ensure they align more effectively with the results and discussion.

Section 1: Hybrid modelling improves performance at local scale.

- The first section "Hybrid modelling improves performance at local scale" doesn't answer the 1st question "How can hybrid process-based and statistical/machine-learning models be effectively integrated to enhance local-scale hydrological predictions across diverse hydrological regimes?" This section of the manuscript simply shows that the 4 post-processing methods used in the study improve the results. This should not be surprising at all, considering that improving the results is the reason one wants to post-process the results in the first place.

Authors Response:

Thank you for this comment. Our goal was to evaluate the performance of the hybrid modelling approach under different evaluation metrics and across different river systems. While the improvements may not be surprising, the comprehensive evaluation across multiple metrics provides new insights. We also analysed which specific aspects are improved and compare the differences between machine learning and statistical methods. We agree that the original question was not well formulated and have revised it, highlighting the focus of representation of different streamflow characteristics.
Actions in the paper:
Line 84: the scientific question is "*(1) How do hybrid process-based and statistical/ML methods enhance local model performance across various streamflow characteristics?*"
Line 121: the corresponding answer is with the title of "*Hybrid modelling improves representation of streamflow characteristics at local scale*", instead of "*Hybrid modelling improves performance at local scale*".

- Most of this section shows the results and explains what the results show, without discussing the details about how the results are unique and why they are important. This section could be improved by further discussing how the authors have done something novel and advanced the science in a way that hasn't been done before. The meaning of the results should also be further discussed.

Authors Response:
We appreciate the reviewer's comment and fully agree with this point. Our novelty lies in the following aspects: (1) a systematic comparison of multiple post-processing approaches, including widely used classical statistical approaches in climate science as well as state-of-the-art machine learning methods; (2) insights gained from evaluation across a wide range of hydrological conditions at a continental scale, (3) and a comprehensive evaluation of the post-processing methods on different streamflow characteristics. In the revised manuscript, we have integrated these points into the Results section.
Actions in the paper:
Line 123: "*We applied four post-processing methods, including both classical statistical and state-of-art ML methods, to correct streamflow simulations across the pan-European region. Our evaluation, focusing on total volume as well as high and low flow extremes, reveals that integrating any of these methods yields substantial improvement in the performance of the underlying process-based E-HYPE model (Fig. 2), while also revealing notable differences in their effectiveness.*"

- Either the question needs to be reframed, or the authors need to indicate how their work answers their first question.

Authors Response:

We agree with the reviewer, and the question has been reframed as suggested in the major comments.

Actions in the paper:

Line 84: the scientific question is "*(1) How do hybrid process-based and statistical/ML methods enhance local model performance across various streamflow characteristics?*"

- I recommend that the discussion for this section conclude with a precise answer for how the authors answer the first question.

Authors Response:

Thank you for this suggestion. We believe this issue has now been addressed by revising both the scientific question and the corresponding subtitle in the Results section. The revised question now is: "How do hybrid process-based and statistical/ML methods enhance local model performance across various streamflow characteristics?" The corresponding answer is now stated in: "Hybrid modelling improves the representation of streamflow characteristics at the local scale." In the Discussion section, we further elaborate on the potential practical application of post-processing methods for improving the accuracy of large-scale hydrological models.

Actions in the paper:

Line 84: the scientific question is now "*(1) How do hybrid process-based and statistical/ML methods enhance local model performance across various streamflow characteristics?*"

Line 121: the corresponding answer is in the section "*Hybrid modelling improves representation of streamflow characteristics at local scale*".

Line 398 in the Discussion section: "*Another key outcome of our study is the potential to produce more accurate forecasts when the hybrid modelling framework is operationalized. This supports global scientific and operational efforts to ensure equitable access to reliable hydrological data, information and services, including the HELPING scientific decade launched by the International Association of Hydrological Sciences (IAHS) [46] and the Early Warnings for All (EW4ALL) initiative launched by the United Nations [47],* *both aiming to protect everyone from hydrological hazards (floods and droughts). To ensure early action to hydrological warnings and long-term planning from decision-makers and stakeholders, accurate hydrological forecasts are essential, which are typically produced and evaluated with reference to model simulations [29,48,49]. However, the biases between model simulations and the observations propagate into the forecast systems [22,43]. These biases can be mitigated by incorporating our proposed hybrid modelling framework, which adjusts the errors inherited from the reference simulation due to imperfection of the hydrological model, improving the forecasts applicability for real-world water resources management and hazard prediction.* ”

- Figure 2 is a sub-figure of Extended Figure 1. Figure 2 can therefore be removed and the text re-written to focus on Extended Figure 1. One of the Chord Diagrams (a sub-figure in Figure 2 or Extended Figure 1) should be more explicitly explained in the methods section.

Authors Response:

We agree with the reviewer for this reorganisation of the figures. We have adjusted the figures accordingly. And added an explanation of interpretation for the chord diagram in the Methodology section.

Actions in the paper:

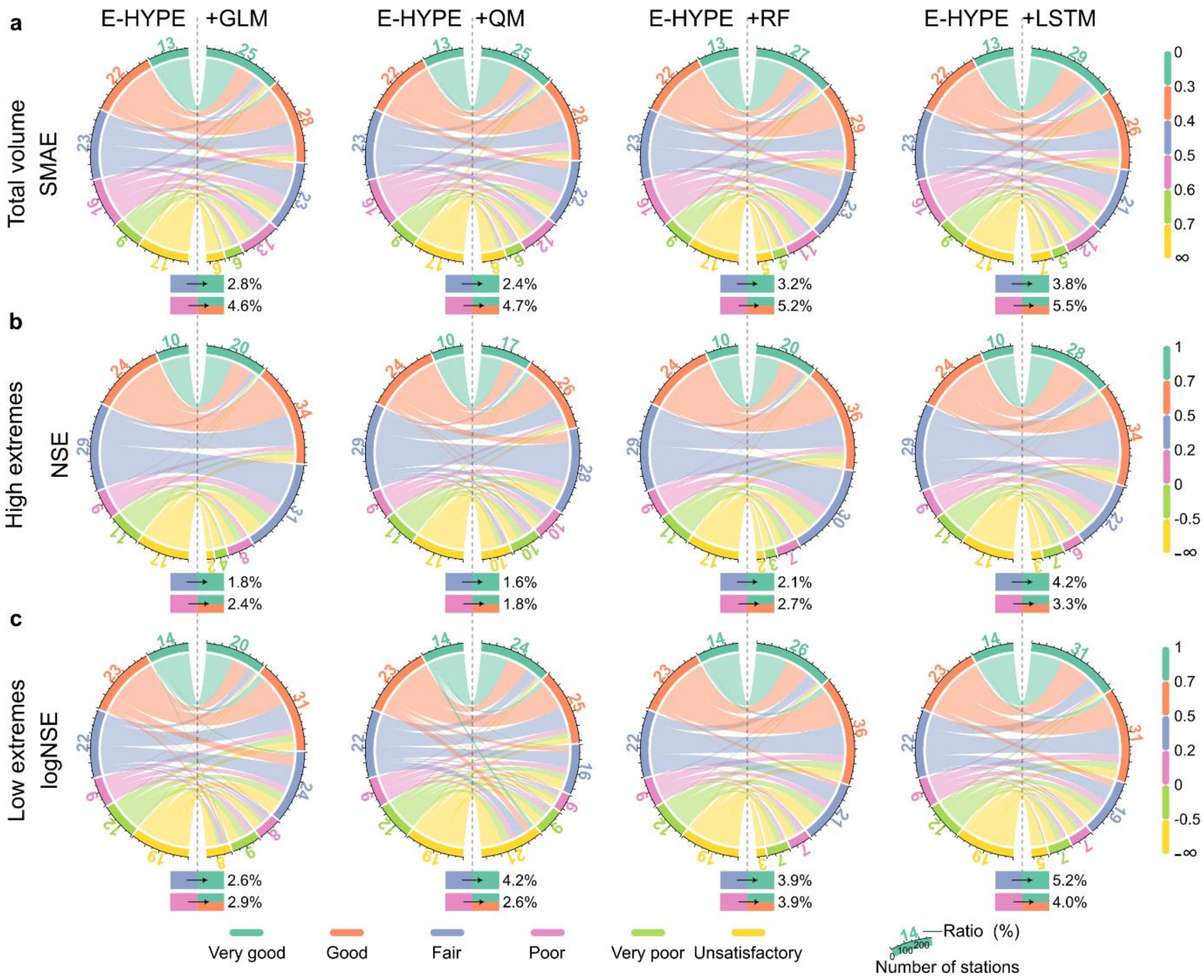


Fig. 3: Chord diagram showing performance transitions before and after post-processing. *The performance for process-based (E-HYPE, left side of the chord) and hybrid models (E-HYPE integrated with GLM, QM, RF, and LSTM, right side of the chord) in predicting streamflow total volume (SMAE; a), high extremes (NSE; b), and low extremes (logNSE; c) are presented. The diagram visualizes how stations transition across six performance groups. The width of each chord represents the proportion of stations shifting between performance groups, highlighting improvements or deteriorations due to post-processing. Portions of stations that experienced significant performance jumps (i.e. from fair to very good, and from poor to good/very good), are displayed below each chord diagram.*

Line 581 in the Methodology section, a description of chord diagram is included: "*The chord diagram (Fig. 3) provides a detailed comparison by tracking the transitions of stations between performance groups before and after post-processing. By depicting the "flow" of stations from one group to another, this visualisation helps clarify the extent to which post-processing methods improve, degrade, or maintain performance across* *different stations. The groups are determined subjectively but still driven by expert knowledge from previous analyses [61]. However, we note that these are not universally applicable [62,63] and are determined specifically for this study.*"

- The terms "very good, good, fair, poor, very poor, and unsatisfactory" shown in Figure 2 and Extended Figure 1 appear to be arbitrary and are not explained in the methods section. What was the criteria used to determine what score corresponds to which term? I would refer the authors to Moriasi et al. (2015) for a published example and Clark et al. (2021) for a critical evaluation of the use of popular performance metrics in hydrologic modeling.

Authors Response:

Thanks for this opinion and the two references. In our study, the primary goal here is to compare performance across methods. We applied a rule-of-thumb approach for the group separation, aiming to maintain a reasonable balance of stations within each group. While this approach may not be universally applicable, we believe it provides sufficient information for our analysis. After reading the references, we recognize that the terminology used for these groups might be open to interpretation. To avoid potential confusion, we have noted in the Methodology section that these groups are specific to our study and may not be directly transferable to other catchments.

Actions in the paper:

Line 584 in the Methodology section: "*The groups are determined subjectively but still driven by expert knowledge from previous analyses [61]. However, we note that these are not universally applicable [62,63] and are determined specifically for this study.*"

Section 2: no single best hybrid method: spatial complementarity of post-processing potential.

- The second section does not answer the question "What mechanisms underpin the spatial complementarity of post-processing methods?" The text describes where each of the methods perform the best without saying very much about what underlying mechanisms explain the results. It's not clear what "mechanisms" the authors are referring to in the question. Are the mechanisms the physical hydrological processes that are dominant in different cases where one post-processing method performs better than the others?

Authors Response:

Thanks for this comment. This confusion is due to the lack of clear correspondence between the scientific questions and the key messages in the Results section. The scientific questions have now been reframed. This section addresses the second question: "How does the performance of different post-processing methods vary across Europe's hydro-climatic gradient?", by comparing the spatial distributions of improvements achieved with each post-processing method and identifying common or distinct patterns. The findings lead to our second key message: "No single best hybrid method: spatial complementarity of post-processing potential".

Actions in the paper:

Line 86: the scientific question is revised as: "*How does the performance of different post-processing methods vary across Europe's hydro-climatic gradient?*"

Line 216: the corresponding answer is in the section with the title of "*No single best hybrid method: spatial complementarity of post-processing potential"*.

- Most of this section shows the results and explains what the results show, without discussing the details about how the results are unique and why they are important. This section could be improved by further discussing how the authors have done something novel and advanced the science in a way that hasn't been done before. The meaning of the results should also be further discussed.

Authors Response:

Thank you for this feedback. We agree that our original text focused only on presenting the results without sufficiently discussing their significance. We have now revised this section by clearly stating that its aim is to evaluate the post-processing methods from spatial aspect, investigate their performance across various hydro-climatic conditions in the pan-European region, and explore whether a universally applicable method exists, in order to bring insight for large-scale hydrological model at a continental scale.

Actions in the paper:

Line 219: "*Building on the overall improvement achieved by the hybrid modelling framework, we now assess its spatial effectiveness, examining how different post-processing methods perform across regions and whether a universally applicable method exists.*"

- Either the question needs to be reframed, or the authors need to indicate how their work answers their second question.

Authors Response:
We agree and we have reframed the scientific question.
Actions in the paper:
Line 86: the scientific question is revised as: "*How does the performance of different post-processing methods vary across Europe's hydro-climatic gradient?*"

- Figure 3 is a sub-component of Extended Figure 2. Figure 3 can therefore be removed and the text re-written to focus on Extended Figure 2.

Authors Response:
As the reviewer suggested, we have removed Figure 3 and incorporated Extended Figure 2 into the revised manuscript, reorganizing its subplots and presenting it as Fig. 4.
Actions in the paper:

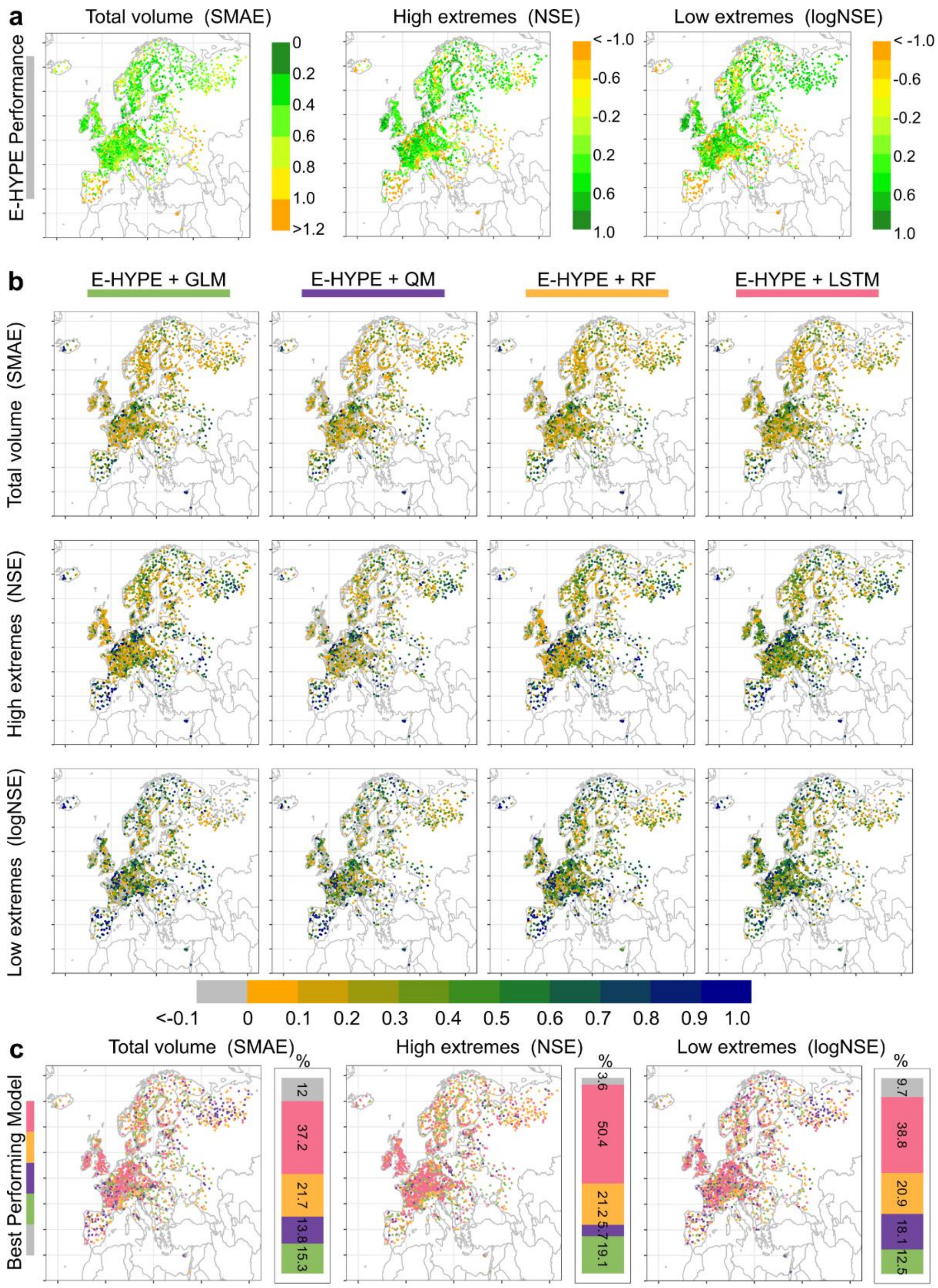


Fig. 4: Spatial distribution of E-HYPE performance, post-processing skill, and best-performing models across different streamflow characteristics. *a, Raw performance of the process-based E-HYPE model. b, Skill improvement from each post-processing method, with colors ranging from yellow to blue indicating performance improvements, while grey represents no improvement after post-processing. c, Best-performing post-processing method at each station and the proportion of stations where each method achieves the highest skill. Colors represent different models: process-based E-HYPE (grey), and hybrid using GLM (green), QM (purple), RF (orange), and LSTM (pink).*

- Extended Data Figure 3 is the first mention of hydrological clusters. I recognize that these clusters are defined in the methods section but should be briefly noted and cited in the results section. Or perhaps summarized and cited in the introduction so that the reader has a sense of what a cluster is when first coming across its use in the text.

Authors Response:

Thank you for this comment, this is an important point. As suggested, we have added an explanation of hydrological clusters in the Introduction and provided a clear reference to the relevant section in the Methodology and the supplementary information.

Actions in the paper:

Line 101: "*However, no single method consistently outperforms across the entire domain, rather a spatial complementarity occurs, which is primarily influenced by catchment characteristics, including hydrological regimes (represented by predefined hydrological clusters[30,32], details in Methodology, Fig. S1 and Table S1 in the supplementary information) and climate conditions among the investigated potential drivers.*"

- Is the "local" portion of the manuscript's title accurate given the statement on lines 187 to 190: "in few river systems, no hybrid method adds value. These stations … being characterized by small upstream area (mostly less than 500 km2)…" The word "local" is used 28 times in the manuscript but is not defined. What exactly is "the local scale" and "local decision making"? The term "local" needs to be precisely defined in the context of the paper. If small upstream areas less than 500 km2 are not local, then what is local?

Authors Response:

We're glad the reviewer brought this up, and we'd like to clarify. By "local scale", we do not mean small catchments but rather the catchment outlets within the context of large-scale hydrological modelling. This refers to the representation of hydrological processes within individual drainage basins, which is usually critical for water management and decision-making. We have added an explanation to clarify this in the text.

Actions in the paper:

Line 36: "*However, LSHMs, especially at national, continental or global levels, face considerable challenges when applied to local scales, referring to locations within the river system which are critical for water management and decision-making [6,7].*"

Section 3: Hydrological regime as a key driver to model performance enhancement.

- Section 3 does not answer the question "How can this inform tailored strategies for optimizing hydrological model performance in different regions?" There is very little discussion around this question with the bulk of the text simply explaining the results.

Authors Response:

Thanks for the feedback. The scientific questions have now been reframed. This section addresses the third question: "What are the key drivers controlling the hybrid model performance enhancement across Europe?", by computing the feature importance of each potential driver and rank them across different post-processing methods. The findings lead to our third key message: "Hydrological regime as a key driver to model performance enhancement".

Actions in the paper:

Line 87, the scientific question is revised as: "*What are the key drivers controlling the hybrid model performance enhancement across Europe?*"

The corresponding answer is in the section with the title of "*Hydrological regime as a key driver to model performance enhancement*" starting from Line 296, with extensive revision.

- The methods section would benefit from more clearly explaining how to interpret the figures in this section.

Authors Response:

We have addressed this comment by adding explanations of the figures in the Methodology section as suggested.

Actions in the paper:

Line 639: "***Visualisations for feature importance***

To assess feature importance, we present results as heatmaps (Fig. 5), where color intensity represents feature importance on a scale from 0 to 100%. The corresponding rankings are visualised as points in the marginal plots, providing a clear comparison of relative feature contributions across the models.

In analysing model skill as a function of key driving factors (Fig. 5; Fig. S3), we illustrate trends using a locally estimated scatterplot smoothing curve (LOESS). This approach captures the general pattern of how model skill varies with numerical driving factors (e.g. temperature or precipitation), providing insight into the underlying relationships. For categorical driving factors as hydrological clusters, model skill is decomposed by cluster group and visualized using violin plots. These plots illustrate the distribution of *skill within each cluster, highlighting variations across hydrological regimes and emphasising the role of river system characteristics in shaping model performance.*"

Section 4: Enabling knowledge transfer for a wider global impact.

- The authors claim in this section that their research is applicable to ungauged locations (lines 280 to 294), but this is not clear at all. Further discussion around this claim is necessary to convince the reader of the claim.

Authors Response:

Thank you for this comment. We have now included an explanation of approaches to integrate the insights from our study for application in ungauged basins. These approaches include: (1) training a regionalized model for basins with similar hydrological characteristics and (2) incorporating hydrological characteristics as input variables to better inform the post-processing model.

Actions in the paper:

Line 379: "*Whilst it is important to note that this connection persists regardless of the post-processing methods employed. Therefore, this new insight provides support for the regionalisation of the post-processors, which is important for the broader application of hydrological models [29,30,43]. Approaches to integrating this insight include training regionalized post-processing models for river systems with similar hydrological characteristics or incorporating these characteristics as input variables to inform post-processing across a broader domain. This enables the transfer of river system-specific knowledge to other geographical locations, even under ungauged conditions. Furthermore, this approach lays a solid scientific foundation for expanding the insights gained from pilot studies to broader applications, particularly in vulnerable areas with limited resources. This aligns with the broader vision that using large-scale hydrological services to identify solutions at the local scale is essential for maximising global impact [44,45].*"

- The authors also appear to claim that their research can support the forecasting of floods and droughts (lines 296 to 308). Again, this is a broad statement that contains very little discussion as to how flow forecasting agencies might be able to use the research. Further discussion is required.

Authors Response:

Thank you for this comment. Our study supports the Early Warning for All initiative in the aspect of enhancing forecast accuracy. This can be achieved by operationalizing the hybrid modelling framework within forecast systems, which help mitigate errors inherited from the process-based model and bring forecasts closer to real-world conditions. We have incorporated this information into the manuscript.

Actions in the paper:

Line 398: "*Another key outcome of our study is the potential to produce more accurate forecasts when the hybrid modelling framework is operationalized. This supports global scientific and operational efforts to ensure equitable access to reliable hydrological data, information and services" …… "However, the biases between model simulations and the observations propagate into the forecast systems [22,43]. These biases can be mitigated by incorporating our proposed hybrid modelling framework, which adjusts the errors inherited from the reference simulation due to imperfection of the hydrological model, improving the forecasts applicability for real-world water resources management and hazard prediction.*"

Methodology

The methodology section is lacking key details.

o A detailed description of all diagrams shown in the analysis would be helpful.

Authors Response:

We have addressed this comment by including detailed explanation for all the diagrams.

Actions in the paper:

From Line 577 in the Methodology section:

"Visualisations for model evaluation

The cumulative distribution plot (Fig. 2) presents the proportion of stations that fall below a given performance threshold for the three evaluation metrics. This allows for an inter-comparison between the different post-processing methods.

The chord diagram (Fig. 3) provides a detailed comparison by tracking the transitions of stations between performance groups before and after post-processing. By depicting the "flow" of stations from one group to another, this visualisation helps clarify the extent to which post-processing methods improve, degrade, or maintain performance across different stations. The groups are determined subjectively but still driven by expert knowledge from previous analyses [61]. However, we note that these are not universally applicable [62,63] and are determined specifically for this study.

To further evaluate model performance across stations, we analyse the spatial distribution of skill by identifying the best-performing method at each station (Fig. 4). The results are visualised using a color-coded map, where each station is assigned a color based on the method that yields the highest performance. Additionally, we calculate the proportion of stations where each method performs best (Fig. 4). These ratios are presented in a bar plot alongside the map, offering a comprehensive view of how different methods perform across the entire study area. This combined visualisation helps highlight spatial patterns in model performance and provides insights into the effectiveness of different post-processing methods."

From Line 638 in the Methodology section:

"***Visualisations for feature importance***

To assess feature importance, we present results as heatmaps (Fig. 5), where color intensity represents feature importance on a scale from 0 to 100%. The corresponding rankings are visualised as points in the marginal plots, providing a clear comparison of relative feature contributions across the models.

In analysing model skill as a function of key driving factors (Fig. 5; Fig. S3), we illustrate trends using a locally estimated scatterplot smoothing curve (LOESS). This approach captures the general pattern of how model skill varies with numerical driving factors (e.g. temperature or precipitation), providing insight into the underlying relationships. For categorical driving factors as hydrological clusters, model skill is decomposed by cluster group and visualized using violin plots. These plots illustrate the distribution of skill within each cluster, highlighting variations across hydrological regimes and emphasising the role of river system characteristics in shaping model performance."

o More details around the post-processing methods would also be helpful.

Authors Response:

We have expanded and restructured this section to provide more details on the post-processing methods, including the training-testing split, target variables, and so on. We also include the corresponding code for the post-processing model training and plotting.

Actions in the paper:

*Line 468: "**Common settings***

The post-processing models take simulated streamflow from the E-HYPE hydrological model as input, with the target variable representing either the observed streamflow (Equation 1) or the relative residual between observed and simulated values (Equation 2). Both observed streamflow and relative residuals were tested as target variables across the methods. An exception is the QM method, which exclusively uses observed streamflow as the target.

$$target_{obs} = y_{obs} \quad \textit{(Equation 1)}$$

$$target_{residual=}(y_{obs} - y_{sim})/(y_{sim} + \varepsilon) \quad \textit{(Equation 2)}$$

where ε is a small constant value introduced to prevent division by zero, particularly in scenarios of low streamflow, ensuring the target variable remains within a reasonable range. By setting the target thresholds to be no smaller than -1, this approach also effectively mitigates the common issue of generating negative streamflow values when using residuals as the target variable.

In the Results section, the target variable yielding the highest performance for each method is selected and presented (Table S2, calculation of the metrics can be found in Table 1), ensuring that the analysis highlights the most effective implementation of each approach.

Each station is corrected independently, with separate model calibration for each location. For model training, the dataset was subsequently divided into training and testing periods, by applying an 80-20% data split. The model evaluation is conducted on the testing periods.

The models are implemented using the R packages randomForest and qmap, along with the Python package TensorFlow. Details on data processing and model training are provided in the code availability section, ensuring reproducibility and transparency.

"

o Regarding the scripts and datasets that were used to perform the analysis. Are the scripts available on GitHub and the data easy to access? This would support reproducibility of the results or expansion of the science to others interested in trying the approaches elsewhere.

Authors Response:

We agree with the reviewer and we have published them openly on Zenodo and provide the link in the data and code availability session.

Actions in the paper:

Line 676: "

***Data availability**: We performed the investigations at the SMHI Hydrology Research unit, where work benefits from joint efforts in developing models and concepts by the whole team. The HYPE model code is available from the HYPEweb portal (https://hypeweb.smhi.se/model-water/). Streamflow data, including E-HYPE model simulations and observations, can be shared upon request, following the SMHI data-sharing policy. Evaluation metrics and model skill assessments can be accessed in the repository on Zenodo [https://zenodo.org/records/14938526].*

***Code availability**: All plots and post-processing models were generated using the R programming language. The code, including scripts for data processing, model training and evaluation, and visualization, can be found in the Zenodo repository [https://zenodo.org/records/14938526]."*

o The Figures are not clearly explained and there is some duplication that can be removed.

Authors Response:

We have addressed this comment carefully and revised for all the figures.

Clark, M., P., Vogel, R., M., Lamontagne, J., R. et al. (2021). The abuse of popular performance metrics in hydrologic modelling. Water Resource Research. 57, e2020WR029001. https://doi.org/10.1029/2020WR029001

Moriasi, D., N., Gitau, M., W., Pai, N., Daggupati, P. (2015). Hydrologic and water quality models: performance measures and evaluation criteria. Transactions of the ASABE. 58(6): 1763-1785. https://doi.org/10.13031/trans.58.10715

Reviewer #2 (Remarks to the Author):

Review report for ´Hybrid approaches enhance hydrological model usability for local streamflow prediction' by Yiheng Du and Ilias G. Pechlivanidis

This manuscript demonstrates how hybrid approaches can improve hydrological simulations through incorporating additional post-processing steps based on machine learning techniques such as random forests, long short-term memory models but more straightforward techniques such as quantile mapping and generalised linear model. While the promising effect of machine learning and hybrid approaches have been shown throughout the field in various studies with different angels and interpretation of hybrid version, this study highlights the potential of including these methods for post-processing at the European scale, which has not been done before. The post-processing approaches were combined with the E-HYPE model, a process-based large-scale hydrological model over the entire European domain. The manuscript, highlights not only the added benefit that the tested methods bring for total volume, high and low flow extremes, but also identifies the potential drivers of the performances, which gives an additional interesting insight to the reader. Overall, this manuscript and it's extensive comparison of over 2000 gauging stations throughout whole Europe and the different post-processing approaches creates an interesting overview of this type of hybrid modelling at a large scale.

Authors Response:
We sincerely appreciate the reviewer's recognition of the significance of our work in large-scale hydrological modelling, and the time and effort the reviewer has invested in this manuscript. We have addressed all comments with our detailed responses below.

While the manuscript is well written, a few comments/suggestions/questions arose:
- Consider adding post-processing in the title to make it more clear to the reader what to expect in terms of hybrid modelling

Authors Response:
Thank you for this comment. We explored various potential titles to incorporate "post-processing" but did not manage to find a satisfying one within the journal's 15-word limit for title. Therefore, we decide to keep the original title, while highlight the importance of post-processing in the abstract and key words.
Actions in the paper:
Line 13: "*Here we explore hybrid methods combining process-based modelling and statistical or machine learning post-processors to improve streamflow predictive accuracy, including extremes, across Europe's hydro-climatic gradient.*"
Line 27: "***Keywords:*** *Hybrid hydrological modelling, Machine learning, Post-processing, Performance attribution, Streamflow prediction*"

- For readability, consider moving the extended data figures (e.g. Extended Data Fig 3, etc.) to the end or appendix if this would not be already the case in the final edit. While these figures include additional information, the current amount of figures is disrupting the reading flow and distracts from the main highlights discussed in the different sections (e.g. Fig 2, Fig 3 and Fig 4b feature importance)

Authors Response:

We agree with the reviewer and have implemented the suggested changes. The plots have been merged and reorganized, and some have been moved to the supplementary information.
Actions in the paper:
The figures have been reorganized as follows: Fig 1 to 5 remain in the main text, while Fig S1 and S2 are settled in the supplementary information.
Fig. 1: Hybrid framework for post-processing process-based LSHM and data availability of the observations.
Fig. 2: Performance comparison of process-based (E-HYPE) and hybrid models (E-HYPE integrated with GLM, QM, RF, and LSTM) in predicting streamflow total volume (SMAE), high extremes (NSE), and low extremes (logNSE).
Fig. 3: Chord diagram showing performance transitions before and after post-processing.
Fig. 4: Spatial distribution of E-HYPE performance, post-processing skill, and best-performing models across different streamflow characteristics.
Fig. 5: Drivers influencing model performance enhancement.
Fig. S1: Properties of available local observations and model-based hydrological clusters.
Fig. S2: Performance and characteristics of stations with no improvement from the hybrid modelling approach in terms of total volume, high and low streamflow extremes.

- Figure 2: it would be interesting to know which stations in which clusters are improving. Or if it’s possible to link it to the clusters? Is there a trend which cluster profits most from the post-processing?

Authors Response:
Thank you for raising this point. We find it very interesting and have incorporated relevant information into Fig. S2 in the supplementary material, where panel (a) shows the proportion of improved and non-improved stations for each hydrological cluster. The observed differences suggest that certain clusters benefit more from post-processing, though this varies depending on the evaluation metric used. Fig. S3 further supports this finding, showing that the clusters with the highest median skills differ for total volume, high flows, and low extremes. Despite these, the clear performance differences across the hydrological clusters remain valid, leading to our third key message: “Hydrological regime as a key driver of model performance enhancement”.
Actions in the paper:

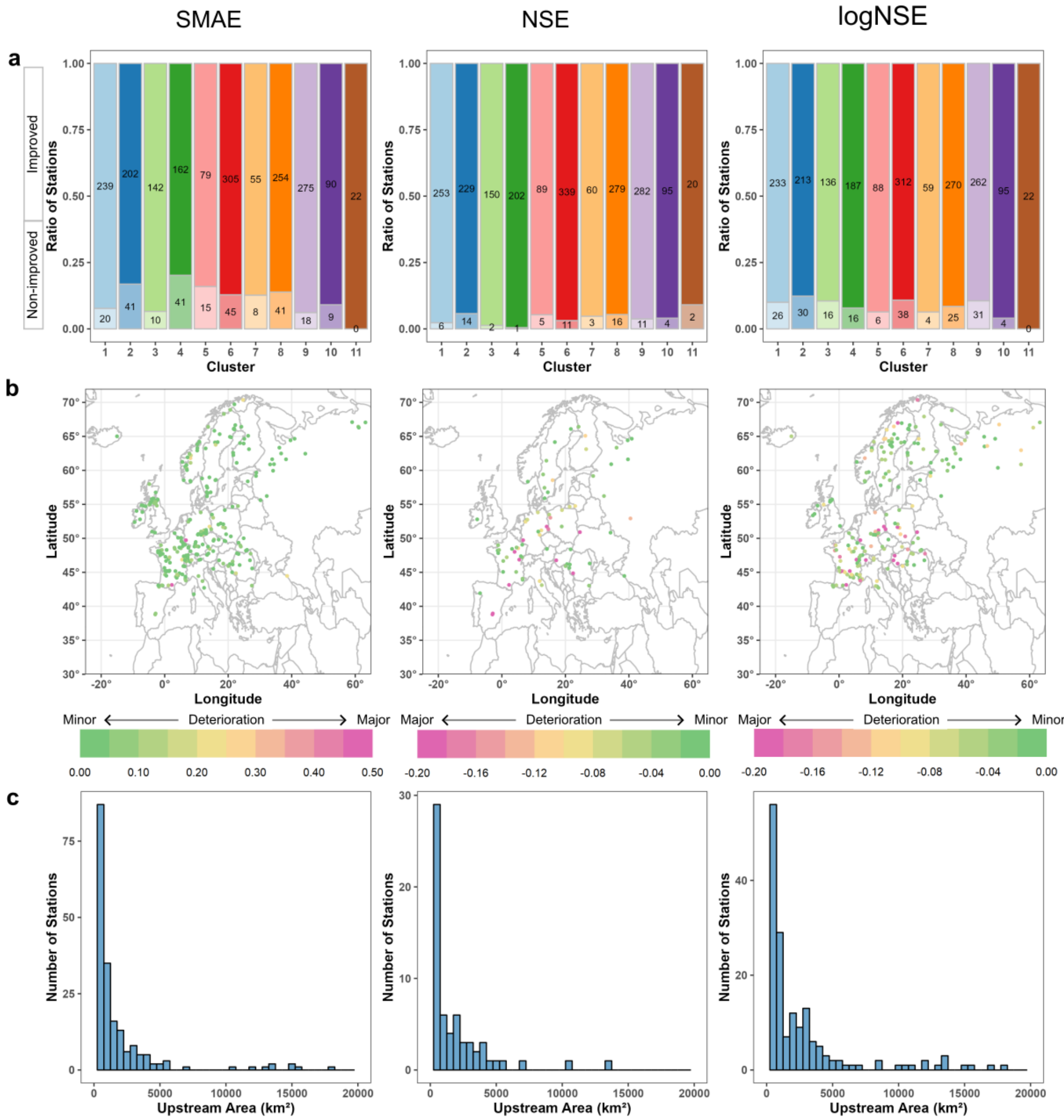


Fig. S2: Performance and characteristics of stations with no improvement from the hybrid modelling approach in terms of total volume, high and low streamflow extremes. a, *ratios of improved/non-improved stations in each predefined hydrological cluster with numbers denoting the actual number of stations. The clusters are used to represent hydrological similarity, detailed information for each cluster can be found in Table S1.* ***b****, performance deterioration in the non-improved stations (performance difference between the process-based E-HYPE model and the best performing hybrid model).* ***c****, frequency distribution of upstream area at the non-improved stations.*

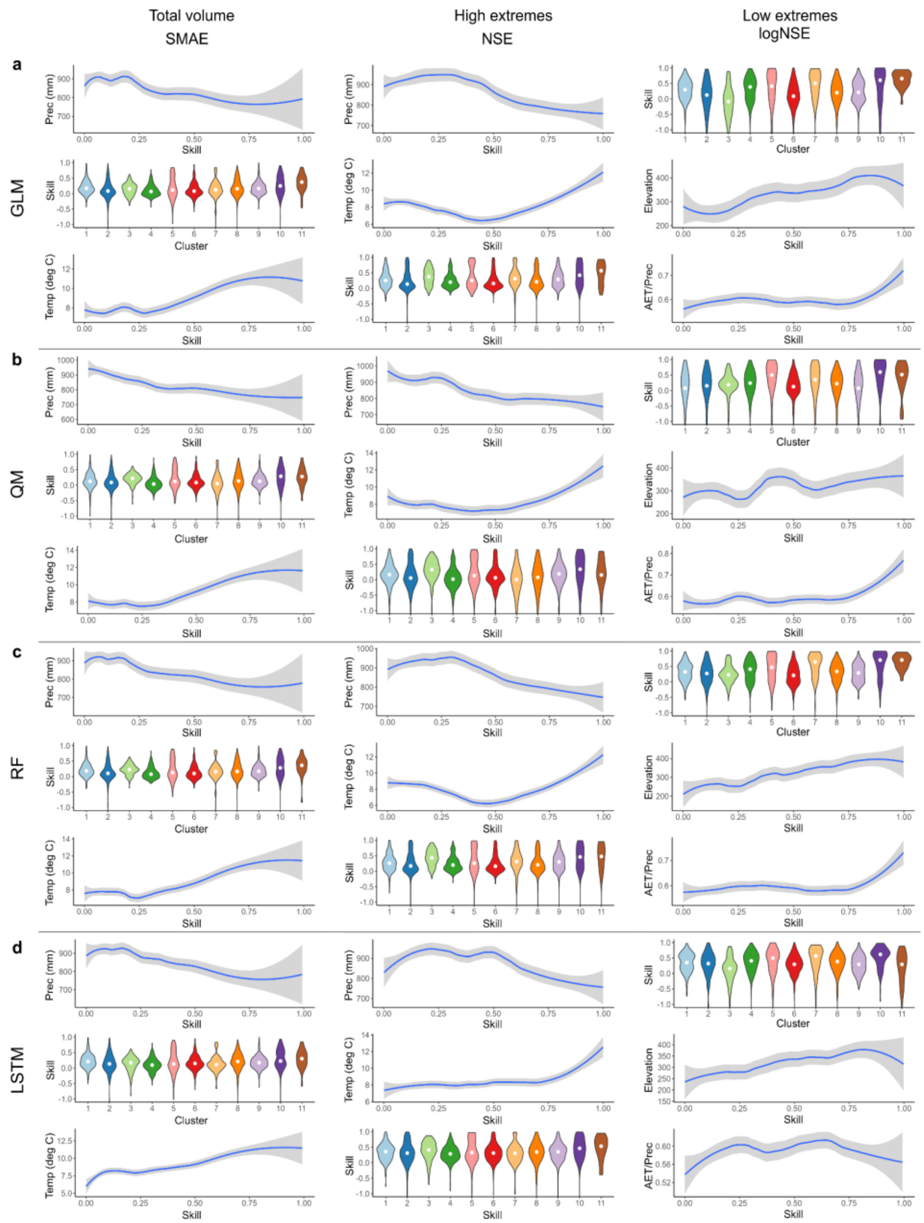


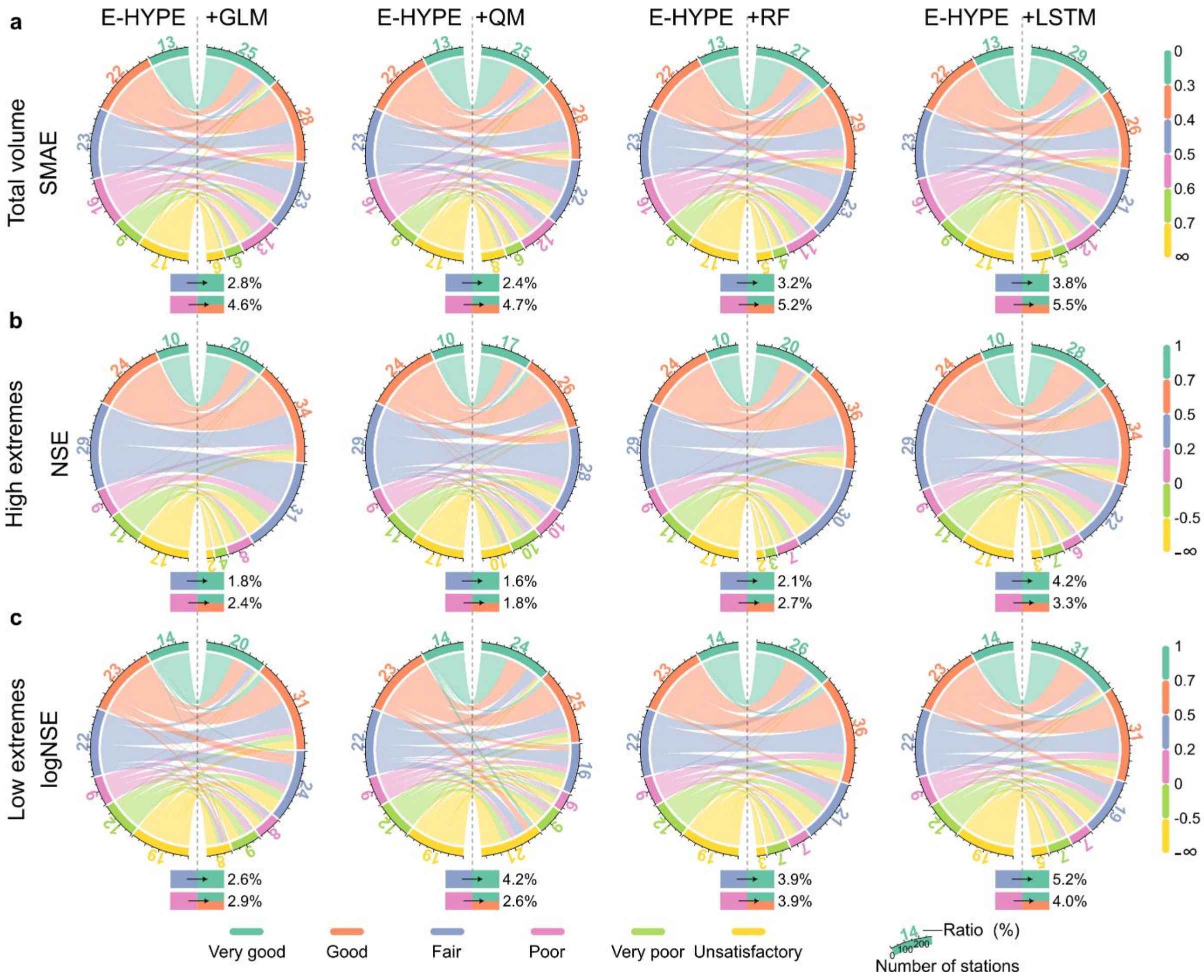
***Fig. S3: Skill variation as a function of the top identified drivers** in terms of total volume - SMAE, high streamflow extreme - NSE, and low streamflow extreme - logNSE for the hybrid models. a, E-HYPE-GLM method; b, E-HYPE-QM method; c, E-HYPE-RF method; d, E-HYPE-LSTM method.*

- Extended Data Fig 1, not better to link it to Figure 2 as Extended Data Figure 2?

Authors Response:

We agree that this was confusing. We have reorganized the figures, and the chord diagram illustrating performance transitions before and after post-processing is now included as Fig. 3.

Actions in the paper:



<u>Fig. 3: Chord diagram showing performance transitions before and after post-processing.</u> *<u>The performance for process-based (E-HYPE, left side of the chord) and hybrid models (E-HYPE integrated with GLM, QM, RF, and LSTM, right side of the chord) in predicting streamflow total volume (SMAE; a), high extremes (NSE; b), and low extremes (logNSE; c) are presented. The diagram visualizes how stations transition across six performance groups. The width of each chord represents the proportion of stations shifting between performance groups, highlighting improvements or deteriorations due to post-processing. Portions of stations that experienced significant performance jumps (i.e. from fair to very good, and from poor to good/very good), are displayed below each chord diagram.</u>*

- Figure 3 nice overview with different post-processing methods

Authors Response:

Thank you! We appreciate this feedback.

- Extended Data Figure 3b) can be confusing if it comes before introducing the clusters. In general, the figure descriptions in the text or captions could be extended in some cases

Authors Response:

We completely agree. We have now introduced the hydrological clusters in the Introduction and included references to detailed information in the Methodology section and supplementary information. Additionally, we have noted this in the figure captions as suggested.

Actions in the paper:

Line 101: "*However, no single method consistently outperforms across the entire domain, rather a spatial complementarity occurs, which is primarily influenced by catchment characteristics, including hydrological regimes (represented by predefined hydrological clusters*[30,32]*, details in Methodology, Fig. S1 and Table S1 in the supplementary information) and climate conditions among the investigated potential drivers."*

"***Fig. S2: Performance and characteristics of stations with no improvement from the hybrid modelling approach in terms of total volume, high and low streamflow extremes. a,*** *ratios of improved/non-improved stations in each predefined hydrological cluster with numbers denoting the actual number of stations. The clusters are used to represent hydrological similarity, detailed information for each cluster can be found in Table S1.* ***b****, performance deterioration in the non-improved stations (performance difference between the process-based E-HYPE model and the best performing hybrid model).* ***c****, frequency distribution of upstream area at the non-improved stations.*"

- Line 232: consider introducing the figure with the clusters from the material and methods also here as the clusters seem to play an important role in the results. Would help the reader follow along better

Authors Response:

We agree and have added reference to the descriptions of hydrological cluster here.

Actions in the paper:

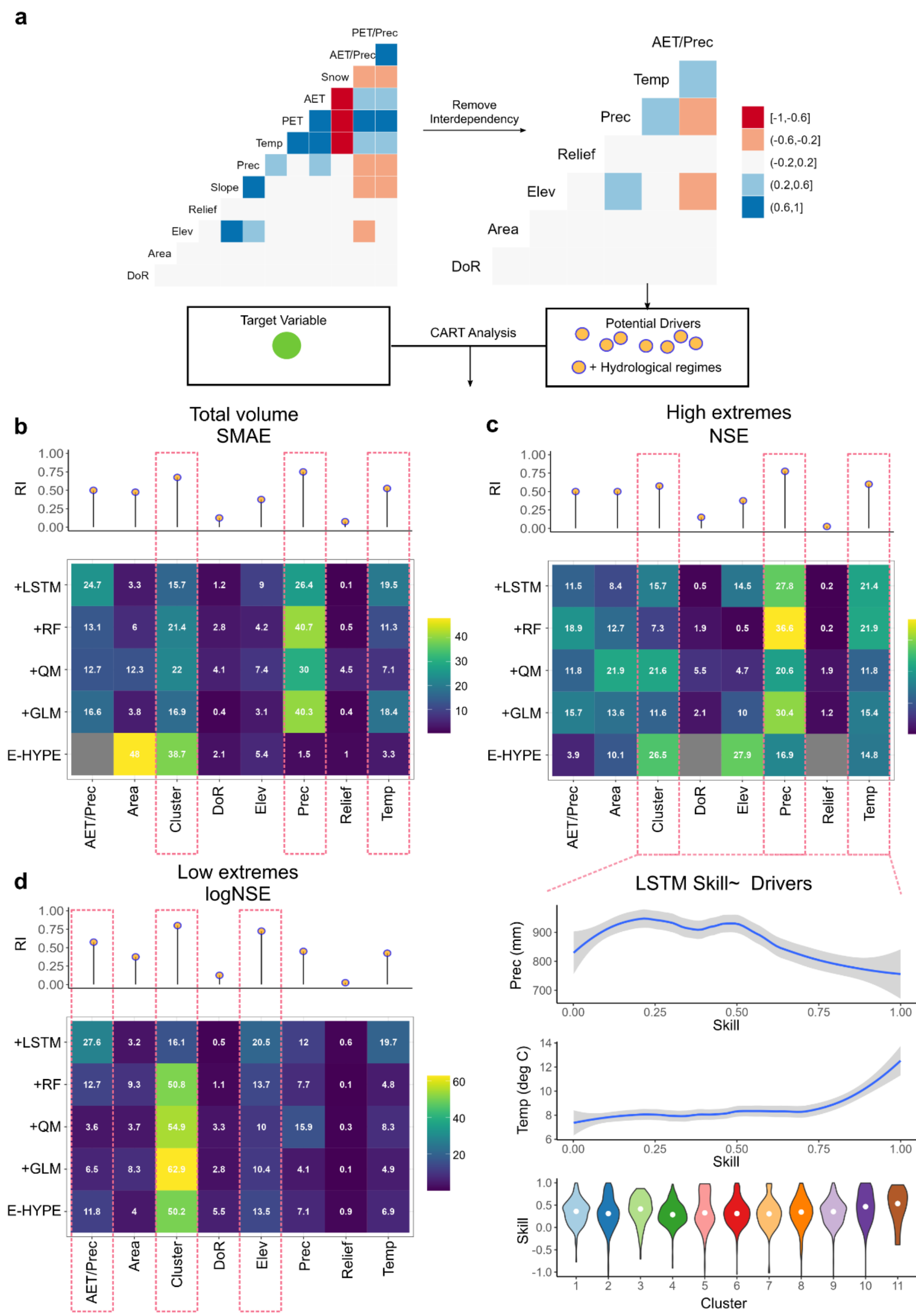
Line 301: "*Here we use predefined clusters of hydrologically similar regimes (Table S1; Fig. S1) across Europe based on a set of hydrological signatures* [30]*.*"

- Figure 4, Extended Data Figure 4 and Figure 5: consider highlighting the feature importance more as these can be easily interpreted on its own or extend the explanation on the interpretation of the skill drivers for easier interpretation

Authors Response:

Thanks for this suggestion. We have reorganized the figures to focus more on feature importance and included in the manuscript as Fig.5. We also added a description in the Methodology section to help with the interpretation of these figures.

Actions in the paper:



[***Fig. 5: Drivers influencing model performance enhancement.*** *a, Potential drivers included and filtered by interdependency (see Table 2 for acronyms). b-d, Feature importance from CART analysis for the process-based and hybrid models for the total volume (SMAE; b), high extremes (NSE; c), and low conditions (log NSE; d). An example of how skill changes as a function of its influencing drivers is also presented (c), showing the skill of E-HYPE-LSTM for high streamflow extremes.*]

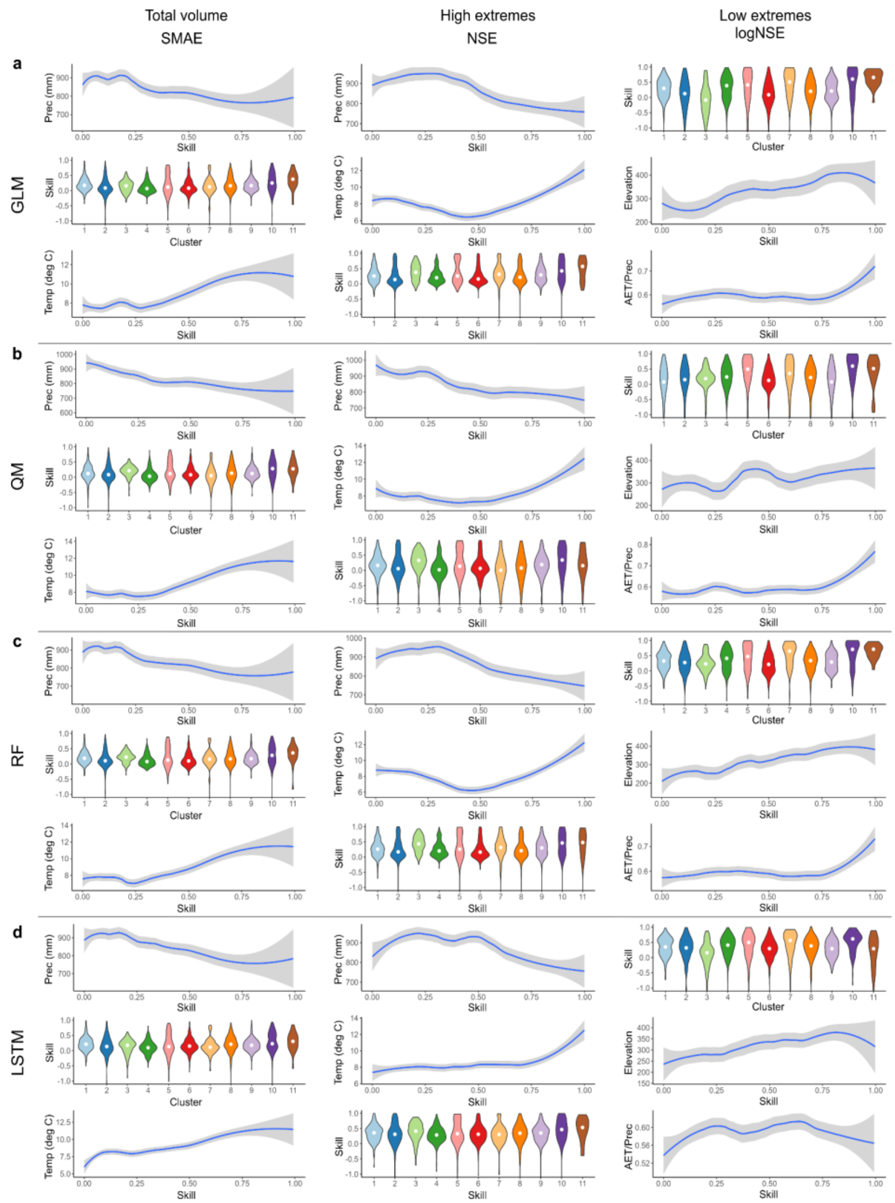
Line 639: "***Visualisations for feature importance***
To assess feature importance, we present results as heatmaps (Fig. 5), where color intensity represents feature importance on a scale from 0 to 100%. The corresponding rankings are visualised as points in the marginal plots, providing a clear comparison of relative feature contributions across the models.
In analysing model skill as a function of key driving factors (Fig. 5; Fig. S3), we illustrate trends using a locally estimated scatterplot smoothing curve (LOESS). This approach captures the general pattern of how model skill varies with numerical driving factors (e.g. temperature or precipitation), providing insight into the underlying relationships. For categorical driving factors as hydrological clusters, model skill is decomposed by cluster group and visualized using violin plots. These plots illustrate the distribution of skill within each cluster, highlighting variations across hydrological regimes and emphasising the role of river system characteristics in shaping model performance."

- Extended Data Figure 5: consider adding total volume, high flow extreme and low flow extreme to the titles of the evaluation scores.

Authors Response:

Thank you for noticing this. This has been addressed and the original Extended Data Fig 5 now is Fig. S3 in the revised version.

Actions in the paper:



[***Fig. S3: Skill variation as a function of the top identified drivers in terms of total volume - SMAE, high streamflow extreme - NSE, and low streamflow extreme - logNSE for the hybrid models.*** *a, E-HYPE-GLM method; b, E-HYPE-QM method; c, E-HYPE-RF method; d, E-HYPE-LSTM method.*]

- Methods: Extended Data Figure 6 would be nice to have in the main body to highlight the flow of the work as it gives a nice overview

Authors Response:

Thank you for this suggestion and we completely agree. We have moved this information to the Introduction and integrated it with the station information in Fig. 1 to provide an overview of the workflow and study area.
Actions in the paper:

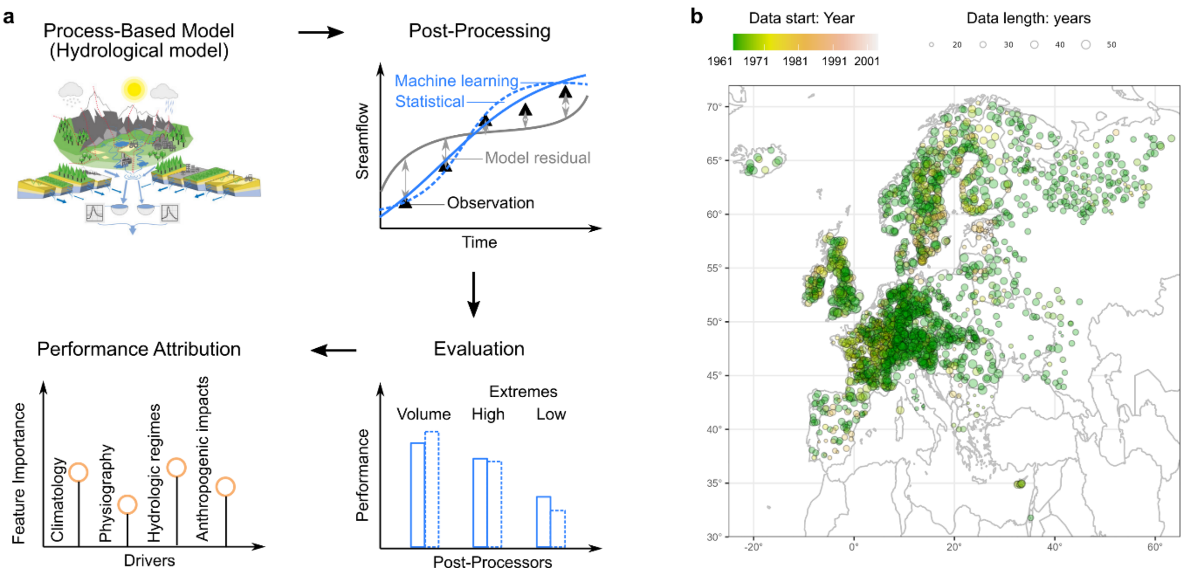


Fig. 1: Hybrid framework for post-processing process-based LSHM and data availability of the observations. *a, Schematic of the hybrid framework, detailing the process-based model, post-processing, evaluation, and performance attribution steps. b, Spatial distribution of the stations used in the study, annotated with the start year and duration of the observational data.*

- Methods: post-processing methods, strongly suggest to highlight that currently it seems that a model per location was developed for the post-processing (not really mentioned otherwise). Furthermore it would be great if the input data for the post-processing step could be clearly stated (currently assuming it's equal to the ones used for the driver analysis?)

Authors Response:
We appreciate that the reviewer raised this point. We have added details in the Methodology section accordingly. Currently, all post-processing methods use only simulated streamflow as input. As the reviewer correctly noted, the post-processing methods are developed for each individual location, meaning that climate conditions and other descriptors are not included since they are unique to each basin. The identified drivers, therefore, are not direct inputs but instead represent underlying

connections influencing post-processing performance from process-understanding aspect.

Actions in the paper:

Line 464, explanation of how post-processing method is trained for each station is added: "*In total four methods were used to post-process the E-HYPE LSHM output to better align with local observations at each individual station…*"

Line 486: "*Each station is corrected independently, with separate model calibration for each location.*"

Line 469, description of input and output variables of the post-processing methods is included:

"The post-processing models take simulated streamflow from the E-HYPE hydrological model as input, with the target variable representing either the observed streamflow (Equation 1) or the relative residual between observed and simulated values (Equation 2). Both observed streamflow and relative residuals were tested as target variables across the methods. An exception is the QM method, which exclusively uses observed streamflow as the target.

$$target_{obs} = y_{obs} \quad \textit{(Equation 1)}$$

$$target_{residual} = (y_{obs} - y_{sim})/(y_{sim} + \varepsilon) \quad \textit{(Equation 2)}$$

where ε is a small constant value introduced to prevent division by zero, particularly in scenarios of low streamflow, ensuring the target variable remains within a reasonable range. By setting the target thresholds to be no smaller than -1, this approach also effectively mitigates the common issue of generating negative streamflow values when using residuals as the target variable.

In the Results section, the target variable yielding the highest performance for each method is selected and presented (Table S2, calculation of the metrics can be found in Table 1), ensuring that the analysis highlights the most effective implementation of each approach."

- Line 391: are the listed steps for training, testing and validation all the same for all the methods?

Authors Response:

Thank you for this comment. The training and testing split are the same for all the post-processing methods. This is now stated in the text.

Actions in the paper:

Line 486: "*Each station is corrected independently, with separate model calibration for each location. For model training, the dataset was subsequently divided into training* *and testing periods, by applying an 80-20% data split. The model evaluation is conducted on the testing periods.*"

- While in the manuscript it was mentioned that the insights can be useful for ungauged areas as well, I was wondering whether the authors also potentially considered making one (e.g. LSTM) per cluster or the whole region? In other studies from the field, there's a discussion that machine learning models trained on various catchments do better for ungauged catchments, which could also be interesting here if the post-processing would be planned to be extended for other locations as well with less observational data for model training.

Authors Response:

We are glad that the reviewer brings up this point. That is indeed our plan, and it is also the purpose of analysing the potential drivers, which will guide us in selecting what variables to include when training a regionalized model. In the revised manuscript, we have added these explanations in the Discussion section. We are currently working on the regionalized post-processing method and have submitted an EGU abstract with the title "Advancing ungauged catchment hydrology through regionalized ML-based post-processing". We look forward to further discussions with the reviewer on this topic at EGU.

Actions in the paper:

Line 382: *"Approaches to integrating this insight include training regionalized post-processing models for river systems with similar hydrological characteristics or incorporating these characteristics as input variables to inform post-processing across a broader domain. This enables the transfer of river system-specific knowledge to other geographical locations, even under ungauged conditions. Furthermore, this approach lays a solid scientific foundation for expanding the insights gained from pilot studies to broader applications, particularly in vulnerable areas with limited resources. This aligns with the broader vision that using large-scale hydrological services to identify solutions at the local scale is essential for maximising global impact [44,45]."*

- Table 3: slope replaced by precipitation?

Authors Response:

Thank you for this question. Here, slope shows a high correlation with mean precipitation, largely due to their common correlation with elevation across the pan-

European region (Fig. 5a). Elevation influences precipitation patterns through orographic effects, leading to condensation and increased precipitation. Steeper slopes can further enhance the orographic lifting, causing air masses to ascend more rapidly, which intensifies precipitation. These together explains the strong correlation between slope and mean precipitation that is observed here.

The following items seems to be the instructions for the reviewer, therefore we have not addressed them.

• What are the major claims of the paper?

• Are the claims novel? If not, please identify the major papers that compromise novelty.

• Will the paper be of interest to others in the field?

• Will the paper influence thinking in the field?

• Are the claims convincing? If not, what further evidence is needed?

• Are there other experiments that would strengthen the paper further? How much would they improve it, and how difficult are they likely to be?

• Are the claims appropriately discussed in the context of previous literature?

• If the manuscript is unacceptable in its present form, does the study seem sufficiently promising that the authors should be encouraged to consider a resubmission in the future

• Is the manuscript clearly written? If not, how could it be made more accessible?

• Could the manuscript be shortened to aid communication of the most important findings?

• Have the authors done themselves justice without overselling their claims?

• Have they been fair in their treatment of previous literature?

• Have they provided sufficient methodological detail that the experiments could be reproduced?

• Is the statistical analysis of the data sound?

• Should the authors be asked to provide further data or methodological information to help others replicate their work? (Such data might include source code for modelling studies, detailed protocols or mathematical derivations).

• Are there any special ethical concerns arising from the use of animals or human subjects?

Reviewer #1 (Remarks to the Author):

I would like to thank the authors for their revision of their paper "Hybrid approaches enhance hydrological model usability for local streamflow prediction." The authors have effectively responded to the majority of my comments as well as the comments of the other reviewer. The paper has, in my view, been improved considerably as a result.
However, it is still unclear to me how the hybrid modelling approach can be used to improve forecasts or to improve predictions in ungauged basins.
Perhaps my lack of clarity stems from not being deeply involved in the hybridization of hydrologic models with machine learning, but I believe that the manuscript could more clearly articulate how such a hybrid approach could be operationalized for someone with my background, which is more focused on hydrologic modelling.
For example, with respect to improving hydrological forecasts, the added text starting on line 408 of the track-change version of the manuscript says that biases between model simulations and observations "can be mitigated by incorporating our proposed modelling framework..." I can understand how this is true for a hind-cast, but not a forecast. How can you adjust errors in a forecast when you don't have observed streamflow (because the future is yet to come)? Are the model parameters somehow changed by the post-processing? Or is the forecasted streamflow somehow changed by the post-processing being applied to the forecasted streamflow without observed streamflow? A few more details would be appreciated.
With respect to ungauged basins, a little more detail is provided, but perhaps not quite enough for someone who isn't doing this exact kind of work on a regular basis. Based on the authors' responses to the other reviewer, it appears that they are working on a regionalization approach for ungauged basins, so perhaps this is worth mentioning in the manuscript along with a few more details.
If the audience of the paper is limited to those who are already heavily involved in the hybridization of hydrologic models with machine learning, then perhaps not much more explanation is needed. For a broader audience, however, clear explanations of how the approach can be operationalized for ungauged basins or forecasting purposes should be included.
If the journal editors wish, I would be happy to review any additional changes made by the authors as this is a very interesting topic of real potential value for operational forecasting agencies around the world.

Authors’ response:
Thank you for your comments. We appreciate the reviewer’s suggestions and fully agree that more detailed explanations on these two aspects will further strengthen the paper’s clarity.

Regarding the operationalization of the hybrid framework, it first needs to be trained using hindcasts (reforecasts) and the corresponding observations. This training is done separately for each lead time, providing lead time-specific correction factors. Once trained, these models can be directly applied to new forecasts, improving accuracy throughout the forecast horizon. We have clarified this process in the revised manuscript.

Regarding the ungauged basins, we are developing a multi-basin post-processing framework that extends the methodology presented in this study. While the current approach trains separate models for each gauged location, the multi-basin framework combines data from multiple basins to train a single model. This model includes static inputs that represent basin-specific characteristics (e.g., climatic conditions, hydrological regimes), allowing it to capture shared hydrological behaviors across basins. As a result, ungauged basins can benefit from the knowledge transferred from hydrologically similar, gauged systems. We have elaborated on this ongoing work in the revised version of the manuscript.

Actions in the manuscript:

We have made the following revisions to the manuscript to address the reviewer's comments in the Discussion section:

Line 311 in the tracked-change version:

"Another key outcome of our study is the potential to produce more accurate forecasts in an operational setting. Hydrological forecast predictability relies on two primary components: the initialization of hydrological conditions at the onset of a forecast, and the hydrological model forcing with (bias-adjusted) meteorological forecasts [30,43]. However, the biases in these two components are inherited from the reference simulation and the quality of the meteorological forecasts which deteriorates as a function of lead time [46,47]. These limitations can both be addressed by training the post-processing models using reforecasts and the corresponding observations for each lead time [48], and consequently providing lead time-specific correction factors. These trained models are then applied to new forecasts, improving accuracy across the entire forecast horizon. Overall, such investigations support global scientific and operational efforts to ensure equitable access to reliable hydrological data, information and services, including the HELPING scientific decade launched by the International Association of Hydrological Sciences (IAHS) [49] and the Early Warnings for All (EW4ALL) initiative launched by the United Nations [50]. Both global community efforts aim to protect everyone from hydrological hazards (floods and droughts), and this achievement relies on accurate hydrological forecasts as a foundation for action, calling for the operationalization of model enhancement efforts."

Line 293 in the tracked-change version:

"A way forward would be the establishment of a multi-basin post-processing approach building on the current method by integrating data from multiple river systems to train a single, regionalized model. While the current framework trains individual models at each gauged basin, the multi-basin approach can incorporate basin characteristics, such as climatic conditions, physiographic attributes, and hydrological regimes, as static input features, enabling the post-processing method to capture shared hydrological behaviours across basins and allowing it to generalize beyond the training locations. Consequently, basins under ungauged or data sparse conditions can benefit from the learned patterns in hydrologically similar gauged systems."