Towards a semi-asynchronous method for hydrological modelling in climate change studies

Frédéric Talbot[1], Simon Ricard[2], Guillaume Drolet[3], Annie Poulin[1], Jean-Luc Martel[1], Richard Arsenault[1], Jean-Daniel Sylvain[1,3]

[1] Hydrology, Climate and Climate Change Laboratory, École de technologie supérieure, Université du Québec, Montréal, H3C 1K3, Canada
[2] Sciences and Engineering, Université Laval, Québec, G1V 0A6, Canada
[3] Direction de la recherche forestière, Ministère des Ressources naturelles et des Forêts, Québec, G1P 3W8, Canada

Correspondence to: Frédéric Talbot (frederic.talbot.2@ens.etsmtl.ca)

RC2 (https://doi.org/10.5194/egusphere-2025-4450-RC2)

General comments:

The authors introduce a novel approach for calibrating hydrological models used in climate change impact studies, and compare the simulation outcomes with the simulation outcomes of two other approaches. The new calibration approach (the semi-asynchronous calibration approach) is based on the existing fully-asynchronous calibration approach but focuses on minimizing differences in simulated and observed discharge distributions for each calendar month, instead of distributions for the full year. The results provide valuable insights and help advance the discussion on calibration methods applied in climate change impact studies.

Reasons for introducing the semi-asynchronous calibration approach are clearly mentioned and the three calibration approaches are extensively compared, by looking at differences in both reference and future periods and by looking at different hydroclimatic variables (not only streamflow). However, clarification of some statements, reconsideration or reflection on certain methodological choices, and more in-depth discussions are needed before the manuscript can be considered for publication.

We sincerely thank the reviewer for the careful and thoughtful evaluation of our manuscript.

The reviewer provides a clear and balanced assessment of the proposed semi-asynchronous calibration approach and of the comparative analysis conducted with the conventional and fully asynchronous methods. We agree that several aspects of the manuscript require clarification, methodological reflection and deeper discussion to strengthen the robustness and practical relevance of the proposed approach.

In the revised manuscript, we propose to carefully address all comments raised by the reviewer. Below, we respond to each specific comment in detail and describe the corresponding revisions we propose to make to the manuscript.

Specific comments:

1. In the abstract and conclusion, it is stated that the semi-asynchronous approach is offered as an compelling alternative for the conventional method (line 32 and line 688). However, lines 519 – 521 state that the semi-asynchronous method does not yet match the stability or reliability of the conventional method. It is suggested to add an explicit practical recommendation on a calibration method for future climate impact studies. In which cases would the semi-asynchronous method be advised?

We thank the reviewer for highlighting this important aspect. We recognize that the manuscript did not provide sufficiently explicit practical recommendations regarding the use of the semi-asynchronous method in climate change impact studies.

Our results indicate that the semi-asynchronous method is particularly suitable in situations where the use of raw climate model outputs is preferred (bias correction can introduce methodological uncertainties, reduce the diversity of climate ensembles, and smooth out extreme events). This approach is also relevant when there is a desire to avoid bias correction, for example, when the goal is to preserve the original climate signal and to better represent extreme flood events, particularly for applications such as infrastructure design under changing climate conditions. In these cases, the semi-asynchronous method maintains the advantages of asynchronous calibration, since it does not rely on bias-corrected inputs, while improving the temporal coherence and seasonal realism that are lacking in the fully asynchronous method.

At the same time, our findings confirm that the conventional method remains the most stable and reliable (inter model variability) approach, especially when hydrological consistency and robustness are the main priorities. For this reason, we do not propose the semi-asynchronous method as a universal replacement for the conventional approach. Instead, we see it as a complementary tool that can be used in parallel with the conventional method.

In practical terms, we recommend that the semi-asynchronous method be used together with the conventional method to better characterize the uncertainty associated with methodological choices in the modelling chain. By comparing results from both approaches, it becomes possible to assess the sensitivity of projected changes to modelling decisions, such as the use of bias correction or direct calibration of climate model outputs.

To clarify these recommendations, we will add a paragraph at the end of the discussion section. This new text explicitly states in which contexts the semi-asynchronous method is advised and explains how it can be used in combination with the conventional method to provide a more comprehensive assessment of uncertainty. Here is the text we propose to add:

"The semi-asynchronous method is especially relevant when there is a need to use raw climate model outputs or when avoiding bias correction is important for preserving the original climate signal and representing extreme events. Although the conventional method remains the most robust and stable option for most applications, the semi-asynchronous approach can be used together with the conventional method to better understand the uncertainty linked to methodological choices."

2. Line 178: "Despite these biases, all models were retained…". A set of 18 climate models was used, despite the large biases for some climate models. It is suggested to reconsider this choice or reflect more extensively on this choice.

We appreciate the reviewer's suggestion to reflect more extensively on the decision to retain all 18 climate models, including those with larger biases. Our intention was to capture a wide range of potential climate outcomes, which is essential for robust climate impact assessments. By including models with varying levels of bias, we were able to evaluate how each calibration method responds to different climate inputs and to assess the sensitivity of the results to the choice of climate model. This approach also allowed us to test the robustness of the fully asynchronous, semi-asynchronous and conventional methods when faced with models that do not reproduce historical patterns as accurately. In the revised manuscript, we will clarify that retaining all models ensures consistency with broader climate impact studies, but also provides a more comprehensive evaluation of the methods under diverse climate conditions. Here is the text we propose to add:

"All models were retained to remain consistent with broader climate impact assessments and to ensure that the ensemble represents a wide range of potential climate outcomes. By including climate models with larger biases, we were also able to evaluate how each calibration method responds to diverse climate inputs and to test the robustness of the approaches under varying conditions. Climate model selection was not performed in this study because the objective was not to develop a credibility-based climate change assessment, but rather to compare the behaviour of the three methods under a common ensemble spanning a wide range of biases. In practical applications, however, screening climate models based on their ability to reproduce key historical climatic features, such as seasonality, may help improve the reliability of hydrological simulations."

3. Lines 249 – 254: Calibration and validation relies on the same observed distribution (26 years). Even though reasons and implications for this choice are described extensively in the manuscript, it is suggested to reconsider this choice. It is expected that an observed discharge distribution based on 13 years can still be a stable and representative reference, particularly because the focus is on the Q5, Q50 and Q95, which are expected to remain approximately when determined for shorter periods. In case a different discharge distribution for the calibration and validation period is used, this is more consistent with the conventional calibration method.

In the asynchronous and semi-asynchronous calibration frameworks, the objective is to match the statistical distribution of observed streamflow rather than to reproduce the year-by-year sequence of historical flows. Climate model outputs do not represent the actual sequence of wet and dry years observed in the historical record, so separating the observed discharge series into two shorter periods would not provide a meaningful or consistent target for calibration and validation.

Using only 13 years to define the observed streamflow distribution would also introduce additional uncertainty. With a shorter period, the distribution of observed flows becomes more sensitive to the presence or absence of specific wet or dry years. This could result in substantial differences between the calibration and validation targets, not because of model performance, but simply due to sampling variability. Since the climate models do not follow the historical sequence of years, there is no mechanism for the models to adapt to these differences, and the comparison would not be robust.

For these reasons, we chose to use the full 26-year observed distribution as a fixed target for both calibration and validation. This approach ensures that the evaluation of the methods is based on a stable and representative reference, and that the results are not unduly influenced by the variability of shorter periods. We believe this choice provides a more consistent and meaningful assessment of model performance within the asynchronous calibration framework.

4. Lines 262 – 266: It is mentioned that calibration and validation periods are separated based on total yearly precipitation from October to September to ensure an equal distribution of wet and dry years between both periods. Calibration is done on 13 years, validation on all 26 years. It is unclear which 13 years are selected; the 13 years with a total yearly precipitation closest to the mean, or the more extreme years? Next to that, please explain why the validation is based on all 26 years instead of the 13 years that are not used for calibration. This would methodologically be more consistent with the conventional method.

We agree that the description of the calibration and validation period selection was not sufficiently clear in the original manuscript. For the calibration period, the 13 years are selected dynamically for each climate model and each catchment in order to ensure a balanced representation of dry and wet years. More specifically, the total precipitation from October to September is first calculated for each year and the years are then ranked from the driest to the wettest. One year out of every two in this ordered series is subsequently selected, which results in a subset of 13 years distributed across the full range of hydroclimatic conditions. These selected years are not necessarily consecutive. During calibration, only the simulated streamflow corresponding to those 13 years is used to evaluate the objective function. We have revised this section of the manuscript to make this procedure clearer.

Regarding the second point, we agree with the reviewer that using the full 26 years for validation may create confusion and is less consistent with a standard split-sample approach. Although this choice was initially motivated by the desire to evaluate model performance against the most robust possible reference distribution, we recognize that a more conventional separation between calibration and validation periods improves methodological clarity and consistency with the conventional method.

For this reason, we will modify the analysis in the revised manuscript and now use a standard split-sample framework with 13 years for calibration and the remaining 13 years for validation. The corresponding text will be updated accordingly, and Table B2 in Appendix B will be revised to present the new validation results.

5. Related to the comment above: In climate impact studies, it is common to do a differential split-sample test to test the performance of the model in climatically contrasting periods. It is recommended to do a differential split-sample test for the conventional method, as this can provide insights into the performance of the model outside calibration conditions.

We agree that evaluating model performance under climatically contrasting conditions is often an important component of climate impact studies. However, a differential split-sample test was not considered appropriate in the present study because it is not compatible with the calibration logic of the asynchronous frameworks. The fully asynchronous and semi-asynchronous methods are calibrated against a fixed streamflow distribution derived from the full observed record, rather than against the temporal sequence of specific years. Calibrating these methods on a subset of climatically contrasting years, such as only dry years, would therefore force the model to identify parameter sets that adapt those years to the full observed distribution. In practice, this could lead to unrealistic parameter adjustments and would likely produce very poor performance when evaluated on wet years, not because of a true lack of model robustness, but because the calibration framework itself is not designed for this type of test.

To maintain methodological consistency and allow a fair comparison among the three approaches, we therefore applied the same split-sample test framework to the conventional method rather than introducing a differential split-sample test for only one calibration approach. In this context, Table B1 presents the calibration and validation results for the conventional method, and these results are also discussed at the beginning of the Results section.

6. Lines 302 – 306: The RMSE value is computed for each calendar month. Please elaborate on the choice for a monthly scale, and not (for example) weakly or seasonal scale. It is suggested to analyse how discharge distributions change when changing from yearly to monthly scale.

This answer is similar to a previous answer to RC1. We agree that the rationale for selecting a monthly temporal scale in the semi-asynchronous method was not sufficiently developed in the original manuscript. The choice of a monthly scale was made to introduce temporal structure into the calibration while preserving the core principle of the asynchronous framework, which is based on matching statistical distributions rather than exact temporal alignment. Moving from a single annual distribution to 12 monthly distributions makes it possible to partially restore temporal coherence and better account for seasonal hydrological processes, while still avoiding full synchronization between observed and simulated time series.

The monthly scale also represents a compromise between coarser and finer temporal discretizations. A seasonal scale would provide less temporal detail and would be less effective at capturing intra-annual variations in hydrological behaviour. Conversely, a weekly or sub-monthly scale would reduce the number of values available within each subset and could make the estimation of discharge distributions less robust. In this sense, monthly discretization was considered an appropriate balance between physical realism and statistical stability.

In addition, monthly aggregation is commonly used in climate and hydrological impact studies, particularly in bias correction approaches such as the conventional framework adopted in this study. Using the same temporal scale therefore also helps maintain methodological consistency across the compared calibration approaches.

We acknowledge that alternative temporal discretizations could be explored and that a more formal analysis of the differences between annual and monthly discharge distributions would further strengthen the methodological discussion. However, such a sensitivity analysis was beyond the scope of the present study. We nevertheless expect that the main conclusions would remain unchanged, as the objective of the monthly formulation is primarily to reintroduce a limited temporal structure rather than to fully redefine the calibration framework.

To address this comment, we will add the following sentence to Section 2.3.4 of the revised manuscript.

"The monthly discretization was selected to capture key seasonal hydrological processes while maintaining enough data points within each subset for robust distribution estimation. Although other temporal aggregations could be considered, such as sub-monthly periods, or moving windows, for example a 30-day window with daily or weekly shifts, they are not expected to alter the main conclusions of this study."

We will also add a sentence at the end of the discussion section to acknowledge this perspective for future work.

"In addition, exploring alternative temporal discretizations, such as sub-monthly groupings, represents a promising avenue for future work to further assess the sensitivity of the method to the chosen temporal structure."

Finally, we note that the comparison between a yearly and a monthly distribution-based formulation is, in essence, already reflected in the present study through the comparison between the fully asynchronous and semi-asynchronous methods.

7. Lines 352 – 357 show RMSE values for the conventional method, but direct comparison with the RMSE values for the asynchronous methods is not entirely appropriate, as the conventional method uses a different objective function during calibration.

We agree that the RMSE values reported for the conventional method are not directly comparable to those obtained for the fully asynchronous and semi-asynchronous methods, since the conventional method was not calibrated using RMSE or a distribution-based objective function.

In this context, it is expected that the asynchronous frameworks perform better according to this metric, as RMSE on discharge distributions is directly aligned with their calibration objective. The purpose of reporting the RMSE values for the conventional method was therefore not to establish a strictly equivalent performance comparison, but rather to provide a reference point and to show how far the conventional method stands from the distribution-based approaches when evaluated using the same diagnostic metric.

To make this clearer, we will add a sentence in the revised manuscript explicitly stating that this comparison should be interpreted with caution and that the better RMSE performance of the asynchronous methods is expected given the differences in calibration objective functions. Here is the sentence we propose to add:

"However, this comparison should be interpreted with caution, as the better RMSE performance of the asynchronous methods is expected given that RMSE on streamflow distributions is directly aligned with their calibration objective, unlike for the conventional method."

8. Lines 378 – 399: Figure 4 shows that the conventional method tends to underperform in reproducing extreme events, mainly high flows. This may relate to the objective function that is selected: KGE for the conventional method, but RMSE for the asynchronous methods. Peak flows dominate the RMSE objective function, while the KGE focuses more on mean flows. It is suggested to check model performance when a model is calibrated using the conventional method but with a similar objective function (RMSE or NSE).

This answer is similar to a previous answer to RC1. To address this point, we recalibrated the conventional method for the Matane catchment using RMSE, in addition to the original calibration based on KGE. This additional analysis was performed to provide a more consistent basis for comparison with the fully asynchronous and semi-asynchronous methods, which both rely on RMSE-based objective functions.

The results obtained for the reference period show that calibrating the conventional method with RMSE does not improve its ability to reproduce high flows compared with the original KGE-based calibration. In fact, the RMSE-based calibration performs slightly worse in this regard. As expected, each calibration performs better according to its own objective function, with the KGE-based calibration yielding a higher KGE and the RMSE-based calibration yielding a lower RMSE. However, both approaches produce very similar hydrographs and overall model behaviour.

These results suggest that, although the choice of objective function introduces some degree of uncertainty, it does not substantially alter the overall performance of the conventional method or the main conclusions of the study. In particular, the lower performance of the conventional method in reproducing extreme flood events cannot be attributed solely to the use of KGE instead of RMSE.

To document this analysis, we added two figures in the response file showing the comparison between the conventional method calibrated with RMSE and KGE for the Matane catchment during the reference period.

To address this point, we will add the following sentence to Section 4.2 and will add the two figures to the Appendix G:

"In addition, the choice of objective function used during calibration of the conventional method represents another source of uncertainty. Nevertheless, sensitivity tests conducted on the Matane catchment showed that using RMSE instead of KGE for the conventional method leads to very similar results, indicating that this choice does not significantly influence the conclusions of the study (Appendix G)."

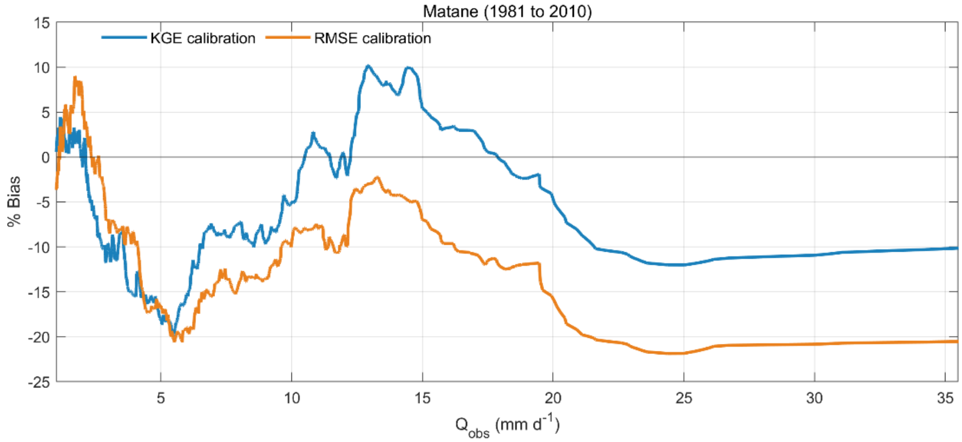

Figure 1. Performance comparison between the conventional method calibrated using RMSE and KGE for the Matane catchment during the reference period (1981–2010). The figure shows the percentage bias between observed and simulated streamflows obtained using meteorological observations.

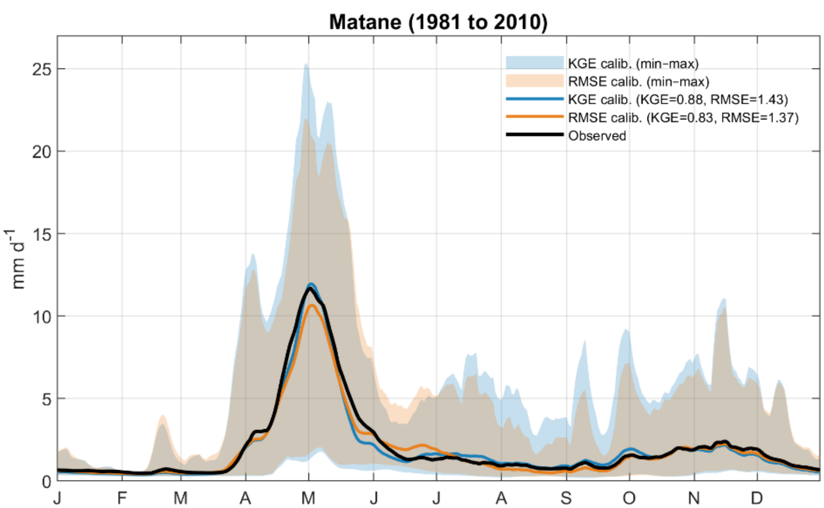

Figure 2. Seasonal streamflow comparison between the conventional method calibrated using RMSE and KGE and observed data across the Matane catchment.

9. Lines 399 – 432: Figure 5(c) shows large biases for modelling low flows, with the semi-asynchronous method having a bias of 95.4% (line 420). This could relate to the choice for the objective function as well. Even though this shortcoming of all three methods is mentioned, a more elaborate discussion on causes and implications of this is missing.

We agree that the original manuscript did not sufficiently discuss the causes and implications of the large biases observed for low flows, particularly for the semi-asynchronous method. One likely explanation is related to the objective functions used in the calibration. Both RMSE and KGE tend to place greater emphasis on medium to high streamflow values, while errors associated with low-flow conditions receive comparatively less weight. As a result, all three methods show more difficulty in accurately reproducing low flows and this limitation is especially visible for the semi-asynchronous method in the reference period.

To clarify this point, we propose to add a sentence in the Results section above Figure 5 to explicitly state that the large low-flow biases likely reflect the lower sensitivity of the objective functions to errors in low-flow conditions.

The following sentence will be added to the revised manuscript.

"These large low-flow biases likely reflect the fact that the objective functions used in this study place greater emphasis on medium and high flows, making all methods less sensitive to errors under low-flow conditions."

10. Line 403 – 426: The absolute mean bias are given, which is the mean of the absolute median bias for each catchment. It is suggested to reconsider using this metric for comparison, as this metric is largely affected by outliers for specific catchments. A median of the absolute median biases could be more appropriate. Additionally, it is suggested to change 'absolute mean bias' to 'mean absolute median bias' to prevent confusion.

We agree that the original terminology and summary metric used in this section could lead to confusion. In the original manuscript, the reference period results were summarized across catchments using the mean of the absolute median biases in order to reflect the overall magnitude of deviations from observations. However, as pointed out by the reviewer, this metric can be sensitive to outliers. Following the reviewer's suggestion, we will revise this part of the analysis and now use the median instead of the mean when summarizing the absolute median biases across catchments. These changes will be incorporated in the revised manuscript and in Tables D1 to D3 of Appendix D.

11. Lines 605 – 621: The issue of equifinality is discussed, but mainly focuses on the lack of the fully-asynchronous method to synchronize certain events/processes. It is suggested to more elaborately discuss how equifinality could be an issue for the semi-asynchronous and conventional method as well. Could it be that calibrating 17 parameters of a physically-based hydrological model lead to overfitting in all three methods? (How) can results be generalized to different types of models, such as more conceptual or empirical models?

We agree that the original discussion focused primarily on the fully asynchronous method and did not sufficiently acknowledge that equifinality can also affect the semi-asynchronous and conventional methods.

In all three calibration approaches, different parameter combinations can potentially produce similar streamflow performance while relying on somewhat different internal process representations. In that sense, equifinality is not exclusive to the fully asynchronous method. However, our results suggest that its magnitude differs substantially among methods. The issue appears much more pronounced for the fully asynchronous method, which is less constrained in time and can, therefore, achieve a good match to streamflow distributions through hydrologically inconsistent combinations of snowmelt timing, surface runoff, interflow, evapotranspiration, and groundwater recharge.

By contrast, the semi-asynchronous and conventional methods are more temporally constrained and therefore more limited in the range of internal process combinations that can produce acceptable streamflow simulations. Although these methods are still subject to equifinality, the problem appears to occur on a much smaller scale, as reflected by the lower intermodel differences observed for internal variables such as snow water equivalent, groundwater recharge and surface runoff. We propose to expand the discussion section accordingly to clarify that equifinality remains relevant for all three methods but is considerably more severe for the fully asynchronous framework.

We do not interpret this issue primarily as overfitting caused by calibrating 17 parameters in a physically based hydrological model. Rather, we interpret it as the existence of multiple parameter combinations that can perform well according to the calibration criterion while still generating somewhat different internal dynamics. This is a general equifinality issue in hydrological modelling, but in our results, it becomes particularly problematic for the fully asynchronous method because those internal differences are much larger and less hydrologically realistic.

Regarding the generalization of the results to other hydrological model structures, we agree that this remains an open question. Since the present study was conducted using a physically based distributed model, it is not possible to draw definitive conclusions for conceptual or empirical models without additional testing. Nevertheless, we do not expect the main conclusions to be specific to this model structure alone. The underlying issue is linked to the degree of temporal constraint imposed by the calibration framework, which should also be relevant for other types of hydrological models. That said, differences in internal process representation may be easier to identify in a physically based model, where internal variables are explicitly simulated and can be evaluated for physical consistency. In simpler conceptual or empirical models, similar effects may still occur, but they could be more difficult to detect or interpret. We therefore acknowledge this as an avenue for future work and it will be clarified it in the discussion by adding this sentence:

"Future work should also assess whether the conclusions drawn in this study hold across different hydrological model structures, including conceptual and empirical models, in order to determine the extent to which the relative behaviour of the conventional, fully asynchronous and semi-asynchronous methods depends on model complexity and process representation."

12. Line 633: "The conventional method (…) shows an increase in intermodal variability in the future…". It is unclear on which result this statement is based. Why would it be a weakness of the conventional method?

We agree that this point was not sufficiently clear in the original manuscript and appreciate the opportunity to clarify it. As shown in Figure 7, the conventional method exhibits relatively low inter-model variability during the reference period compared to the asynchronous and semi-asynchronous methods. This reduced variability is primarily due to the bias correction step, which tends to constrain the climate model outputs toward observed conditions.

However, in the future period, the inter-model variability of the conventional method increases and becomes comparable to that of the semi-asynchronous method for several variables (e.g., streamflow, interflow, groundwater recharge, and SWE).

This increase arises because bias correction does not constrain the magnitude of projected changes and differences between climate models in their future projections are therefore more directly reflected in the simulations. This is an expected behaviour, as we expect climate models' variability to increase in future periods.

In contrast, the variability of the asynchronous and semi-asynchronous methods remains more stable between the reference and future periods, as these approaches directly incorporate the raw climate model signals during calibration. This behaviour is not necessarily a weakness of the conventional method, but rather a characteristic of how it processes climate data. This does not reduce ensemble diversity per se, but it changes how variability is distributed between periods.

To improve clarity, this explanation will be explicitly incorporated into the revised manuscript, both in the description of Figure 7 and in Section 4.1.

Technical corrections:

1. Lines 13 – 17 & Lines 90 – 92: In the abstract and introduction, it is suggested to describe the semi-asynchronous calibration approach. Rather than mentioning 'incorporating a monthly temporal structure', it could help to explicitly mention 'calibration on monthly observed discharge distributions'.

This wording will be revised in the abstract and introduction.

2. Line 76: 'Despite these challenges, they recommend using a pre-processing approach rather than post-processing for climate impact studies.'. This sentence may confuse the reader, as in the end the proposed semi-asynchronous calibration method uses post-processing.

We agree with the reviewer that this sentence could be confusing in the context of the fully proposed and semi-asynchronous methods. It will therefore be removed from the revised manuscript.

3. Line 82: 'Albeit at the cost of reduced temporal synchronicity'. There is not temporal synchronicity at all for the fully-asynchronous approach. Suggested to remove 'reduced'.

The changes will be made to the revised manuscript.

4. Line 138: Information is given on the temporal resolution of the discharge data (daily measurements), but information on the spatial resolution is lacking. Is the discharge data available for the outlet point of the catchment only, or are multiple points in the catchment considered?

Streamflow data were available only at one hydrometric station per catchment, located at the outlet. This clarification will be added to the revised manuscript.

5. Line 151: 'Established relationships'. Suggested to add a reference.

A reference will be added in the revised manuscript.

6. Lines 233: ‘Climate change studies were subsequently conducted using bias-corrected climate model data’. Suggested to change ‘studies’ to ‘simulations’.

This sentence will be revised in the manuscript.

7. Line 295: The added value of the workflow diagram at the end of section 2.3.3 to the manuscript is unclear. Suggested to remove the diagram and only refer to Ricard et al. (2023), or consider moving the diagram to the beginning of section 2.3.3.

We will move the diagram to the beginning of section 2.3.3.

8. Lines 344 – 357: Consider summarizing the KGE and RMSE values in a table for a quicker and more concise overview.

A summary of these results is provided in the main text, while the detailed values remain in Appendix B, as they are not central to the main research question.

9. Line 352 – 353: Units are missing.

Units will be added to the revised manuscript.

10. Line 460: Figure 6 is unclear. The explanation of the legend items ‘reference’, ‘future’ is missing. It is unclear if the bars and lines belong to precipitation or temperature or vice versa. Figure caption can be improved.

The caption will be revised in the manuscript.

11. Line 555: Figure 9. Units are missing.

Figure 9 shows relative soil moisture, which is a unitless variable. To make this clearer, we will add the word “relative” to the figure caption in the revised manuscript.

12. Line 586: ‘(…) irrespective of the methodological differences in how streamflow is simulated’. For both the fully-asynchronous and semi-asynchronous method, streamflow is simulated similarly (using the WaSiM model), but the calibration approach was different. Reconsider the formulation of this sentence.

We agree that the original formulation was imprecise, since streamflow is simulated using the same hydrological model in all cases, while the methodological differences lie in the calibration approach. The sentence will therefore be revised by replacing “*irrespective of the methodological differences in how streamflow is simulated*” with “*irrespective of the methodological differences in the model calibration*” in the revised manuscript.

13. In general, some sections of the manuscript could be written more concisely. Some suggestions:
    - Lines 27 – 30 are mostly a repetition of lines 20 – 24.

This repeated passage will be removed from the revised manuscript.

- Lines 114 – 119 are a repetition of lines 109 – 110.

This repeated passage will be removed from the revised manuscript.

- Lines 189 – 191 are a repetition of lines 186 – 187.

Lines 189 to 191 will be removed in the revised manuscript to avoid repetition.

- Reconsider to remove the sentence in lines 313 - 315, as it does not add new information to the manuscript.

We agree that this sentence did not add substantial information and it will be removed from the revised manuscript.

- Lines 486 – 488: 'The left column…' Consider adding these kind of descriptions in the table and figure captions instead of in the text for better readability.

The description previously included in lines 486 to 488 will be moved to the caption of Figure 7. The same approach will be applied throughout the manuscript to improve readability.

- Section 4.1 is quite long. Consider writing this paragraph more concisely or add sub-headings for readability.

To improve readability, the former Section 4.1 will be divided into two sections. Section 4.1 remains focused on hydroclimatic variable representation, while a new Section 4.2 on equifinality will be added. As a result, the former Section 4.2 on limitations and future directions will become Section 4.3.

- Lines 691 – 698: This paragraph can be shortened or removed for conciseness.

This paragraph will be removed from the revised manuscript for conciseness.

14. Reference list:

- Line 972: A doi link is missing.

The missing DOI link will be added to the reference list.

- Line 988: The reference is to a pre-print, while a peer-reviewed final revised version is available as well.

The reference will be updated to cite the final peer-reviewed version instead of the preprint.

We sincerely thank the reviewer for the careful reading of our manuscript and for the constructive comments provided. These suggestions helped us improve the clarity, conciseness and overall quality of the revised version.

Sincerely,

Frédéric Talbot on behalf of all authors