

Dear editor and reviewers,

Thank you for your thorough review of our manuscript. We sincerely appreciate your evaluation and the constructive suggestions you provided, which certainly helped us improve the quality of our study. We have carefully considered all of your comments and revised the manuscript. The following are point-to-point response to each comment including all relevant changes made in the manuscript.

Editor

Comment 1: Three reviewers have raised several concerns and comments on how to improve this manuscript. For me, the consistency of the model's temporal scale in model output and calibration, in-depth analysis of streamflow variation with different streamflow indicators and lack of physical interpretation of the outcomes should be carefully addressed.

Response: Thank you for your comments and suggestions, as well as those from the three reviewers. Detailed responses to each of the reviewers' comments are provided below. Regarding your first concern about the inconsistency between the model output and calibration time scales, I would like to offer the following explanation. The purpose of this paper is to assess the value of data with different temporal resolutions in hydrological modeling. The computational time step of the model could be a factor influencing model performance, and we wanted to exclude this factor to focus on our research question. Therefore, in each test case, we set the model's computational time step uniformly to 1 hour, resulting in an original model output at an hourly scale. On the other hand, the value of observed streamflow data to hydrological models lies in its use as a benchmark for parameter calibration. This requires aggregating the model output to various time scales and comparing it with observed data at different resolutions to calibrate model parameters. Thus, when accounting for both the need to eliminate the impact of computational time step and the value of observed streamflow data at different resolutions, the inconsistency between model output and calibration time scale becomes unavoidable. We have made some revisions in Section 2.3, "Model experimental design," to clarify the reason for this inconsistency.

To address the need for more in-depth analysis and physical interpretation of the outcomes, and in line with your and the other reviewers' comments, we have selected three representative catchments and provided a detailed description in Section 3.4 of the revised manuscript. These three representative catchments were selected based on a comprehensive consideration of basin attributes including drainage area, rainfall gauging area and streamflow variability, which were identified as the most important factors impacting the model performance and its improvement. Through specific case studies, we demonstrate that the ability of increased data resolution to introduce new, accurate, and effective information is crucial in determining its potential value to enhance model performance. The GOUE index employed in this study offers a partial indication of the potential informational content that high-resolution data may contribute.

In the discussion section, in conjunction with other existing studies, we also presented several additional reasons that may limit the value of high-resolution data, including the lower signal-to-noise ratio of high-resolution data, the model's ability and suitability to capture high-resolution data, and the simplifications of real hydrological processes within the model.

Reviewer #1

Comment 1: In addition to the rainfall and runoff data already mentioned in the paper, other data used for modeling should also be detailed.

Response: Thank you for your reminder. Indeed, we also used DEM, potential evapotranspiration, Temperature, Normalized Difference Vegetation Index (NDVI) and Leaf Area Index (LAI) data necessary to drive the THREW model in this study. The DEM in this study was from the MERIT Digital Elevation Model (DEM) with a spatial resolution of 90m. Temperature and potential evapotranspiration data were sourced from ERA5-land. NDVI and LAI were obtained from MODIS products. These descriptions have been added to the last paragraph of Section 2.1.

Comment 2: It would be beneficial to reference Figure 2 at the beginning of Section 2.3 when introducing the experimental design. This will help readers refer to the flowchart while reading the corresponding text, improving their understanding of the experimental setup.

Response: Thanks for your suggestion. We have referred to Figure 2 at the beginning of Section 2.3 to enhance the reader's understanding of our experimental setup.

Comment 3: In the hourly tests, the model's original output was at the hourly scale, while the calibration was conducted at different time scales, and then the evaluation metrics were calculated back at the hourly scale. Why? I believe it is necessary to provide further clarification on the rationale behind this design choice.

Response: As the primary approach of this study, the experimental design indeed needs to be clearly and thoroughly explained. We apologize for not making it clear earlier. I would like to first explain the rationale behind this design:

1) Daily and hourly scales are the most concerned and commonly used scales in current hydrological modeling. The purposes of the daily and hourly tests are to investigate the effects of input data and measured streamflow with different resolutions on daily streamflow and hourly streamflow simulations, respectively. Therefore, the final evaluation metrics are the assessments of the simulated daily and hourly streamflow results.

2) To eliminate potential impacts of the model's computational time step, the time step was standardized to 1 hour across all cases in both tests. This led to the model's original output being at the hourly scale in all cases. In the hourly test, the observed streamflow was also treated as input data for model to its calibration, and its resolution was the subject of investigation. Therefore, to explore the effects of observed streamflow data at different resolutions, the model's original hourly simulated streamflow was aggregated into different time steps, which were then used for model calibration.

In Section 2.3 in the revised manuscript, when introducing the two experiments, we first clarified our research objectives, and appropriately added some detailed descriptions to ensure that readers can clearly understand the rationale behind our experimental design.

Comment 4: In the results shown in Figure 3, the REP values are negative for most watersheds in both the daily and hourly tests. Does this mean there is a systematic underestimation of peak values in the model?

Response: Accurately simulating peak flows is a challenge for all models. This is especially true for small and medium-sized basins like those in this study, where streamflow changes relatively quickly and flood durations are shorter, making it more difficult to accurately model peak flows. Results in this study also indicate that the THREW model performs better in basins with larger areas, higher precipitation, and smaller diurnal runoff variations.

In this study, peak flows were underestimated in most basins, which might be due to the use of the single-objective KGE metric during model calibration. The KGE metric reflects the model's accuracy over the entire time series. The model is a simplification of natural runoff processes and operates with fixed parameters throughout its execution. The accuracy of the model's simulation of peak flows and other processes (as evaluated by REP and KGE) are sometimes conflicting objectives. Therefore, to achieve a higher KGE, the optimization algorithm tends to select parameters that improve the overall simulation accuracy, which leads to the underestimation of peak flows in most small basins.

This study focuses on the performance trends of the model under different data resolution conditions. All experiments strictly adhered to the principle of controlling variables. We used the same model, parameter range, and parameter optimization algorithm throughout the study, and the model performed well overall. Therefore, even if there was a systematic underestimation of peak values, the trend remains consistent. Therefore, the conclusions of this study are not affected by the systematic underestimation.

Comment 5: The paper currently concludes that higher resolution data does not necessarily improve prediction accuracy. It would strengthen the paper to include further analysis or discussion from a hydrological mechanism perspective to explore the underlying reasons for this finding.

Response: Thanks for your suggestion. Other reviewers have made the same recommendations, and we recognize the need for a deeper analysis. Therefore, combined with suggestions from other reviewers, we have selected three representative catchments based on a comprehensive consideration of basin attributes including drainage area, rainfall gauging area and streamflow variability, which were identified as the most important factors impacting the model performance and its improvement. Hourly time series of rainfall, simulated flow and observed flow under 1-hour and 24-hour resolution data for the representative catchments and representative flood events have been presented in Section 3.4 of the revised manuscript.

Through the time series of observed and simulated flow at the representative catchments, this study intuitively showed that in smaller catchments with larger streamflow variations (e.g., Catchment 1), 24-hour resolution data struggles to accurately capture the true rainfall-runoff processes, and even fails to distinguish between different events. In such catchments, increasing the data resolution provides enough additional effective information, which significantly enhances the accuracy of simulations. In catchments with moderate area and diurnal streamflow variations (e.g., Catchment 2), the difference between 24-hour resolution data and observed values is relatively large, particularly for peak flows. However, this difference is markedly reduced when the data resolution is increased to 6 hours. Further increases in data resolution may not yield additional effective information, resulting in limited improvement in model accuracy. In catchments with large area and small diurnal streamflow variations (e.g., Catchment 3), the observed flow at a 24-hour resolution already closely approximates actual flow, with the relative error in peak flow being only -8.8%, compared to -62.3% and -28.9% for the peak flow errors in the other two catchments. The discrepancy is even negligible when the resolution is increased to 6-hour. Thus, when using 24-hour resolution data, the model performs well on an hourly scale, and continuous improvement in resolution did not lead to a notable increase in simulation accuracy.

In summary, from a mechanistic perspective, this study identified drainage area, rainfall gauging area and streamflow variability as key factors influencing the value of high-resolution data. Through specific case studies, we demonstrate that the ability of increased data resolution to introduce new, accurate, and effective information is crucial in determining its potential value to enhance model performance. The GOUE index employed in this study offers a partial indication of the potential informational content that high-resolution data may contribute.

In the discussion section, in conjunction with other existing studies, we also presented several additional reasons that may limit the value of high-resolution data, including the lower signal-to-noise ratio of high-resolution data, the model's ability and suitability to capture high-resolution data, and the simplifications of real hydrological processes within the model.

Reviewer #2

Comment 1: My biggest concern is this work could be better instead of just a labor work. We are wondering the differences between the simulated streamflow using higher and lower resolutions of rainfall and streamflow, e.g., the timing and the magnitude etc. At least authors should plot the streamflow time series at several representative gauges to be transparent about the results. Sometimes, the metrics cannot fully tell the true performance or even give wrong indications.

Response: Thank you for your constructive suggestion. We recognized the necessity of presenting representative data and conducting a more in-depth analysis. Therefore, we have selected three representative catchments based on a comprehensive consideration of catchment's attributes such as drainage area, rainfall gauge area, and streamflow variability, which were identified as the most important factors impacting the model performance and its improvement. Hourly time series of rainfall, simulated flow and observed flow under 1-hour and 24-hour resolution data for the representative catchments and representative flood events have been presented in Section 3.4 of the revised manuscript.

Through the time series of observed and simulated flow at the representative catchments, this study intuitively showed that in smaller catchments with larger streamflow variations (e.g., Catchment 1), 24-hour resolution data struggles to accurately capture the true rainfall-runoff processes, and even fails to distinguish between different events. In such catchments, increasing the data resolution provides enough additional effective information, which significantly enhances the accuracy of simulations. In catchments with moderate area and diurnal streamflow variations (e.g., Catchment 2), the difference between 24-hour resolution data and observed values is relatively large, particularly for peak flows. However, this difference is markedly reduced when the data resolution is increased to 6 hours. Further increases in data resolution may not yield additional effective information, resulting in limited improvement in model accuracy. In catchments with large area and small diurnal streamflow variations (e.g., Catchment 3), the observed flow at a 24-hour resolution already closely approximates actual flow, with the relative error in peak flow being only -8.8%, compared to -62.3% and -28.9% for the peak flow errors in the other two catchments. The discrepancy is even negligible when the resolution is increased to 6-hour. Thus, when using 24-hour resolution data, the model performs well on an hourly scale, and continuous improvement in resolution did not lead to a notable increase in simulation accuracy.

In summary, through specific case studies, it was demonstrated that the ability of increased data resolution to introduce new, accurate, and effective information is crucial in determining its potential value to enhance model performance. The GOUE index employed in this study offers a partial indication of the potential informational content that high-resolution data may contribute.

Comment 2: I also concern the model used and the study area, which determine the generality of the conclusions, though authors acknowledge such limitations in the manuscript.

Response: The generality of the conclusions to other regions and models is also a concern for us and other researchers. As we mentioned in the Discussion and Limitations sections, other authors (e.g., Ficchi et al., 2016; Reynolds et al, 2017) have reached similar conclusions in related studies, and they also mentioned concerns about generality and the need for further research. In order to verify the generality of this conclusion across different climatic zones and models, we deliberately conducted similar research in northern China using another model (this work is currently under discussion at egusphere-2024-2966), and obtained similar conclusions. In that work in northern China, we also found that for daily streamflow simulations, improvements in model performance become negligible when the resolution exceeds 12 hours. As for hourly streamflow simulations, improvements in overall flood process accuracy become negligible when the resolution of input exceeds 6 hours, while higher resolutions further enhance the precision of peak flow simulations. We have addressed this issue in the Discussion section of revised manuscript.

We also anticipate the advent of further scholarly investigations encompassing diverse geographical areas and model frameworks. We believe that as more related studies emerge, we will gain a clearer and more comprehensive understanding.

References:

Ficchi, A., Perrin, C., and Andreassian, V.: Impact of temporal resolution of inputs on hydrological model performance: An analysis based on 2400 flood events, *Journal of Hydrology*, 538, 454-470, 10.1016/j.jhydrol.2016.04.016, 2016.

Reynolds, J. E., Halldin, S., Xu, C. Y., Seibert, J., and Kauffeldt, A.: Sub-daily runoff predictions using parameters calibrated on the basis of data with a daily temporal resolution, *Journal of Hydrology*, 550, 399-411, 10.1016/j.jhydrol.2017.05.012, 2017.

Tudaji, M., Nan, Y., and Tian, F.: Assessing the Value of High-Resolution Data and Parameters Transferability Across Temporal Scales in Hydrological Modeling: A Case Study in Northern China, *EGUphere* [preprint], <https://doi.org/10.5194/egusphere-2024-2966>, 2024.

Comment 3: Line23: runoff is not input.

Response: We apologize for the confusion caused by our wording. Our intention was to emphasize the importance of rainfall and observed streamflow data to the model. We revised the original sentence in the manuscript as follows:

"The effectiveness of these models heavily depends on the quality and resolution of the data, especially the rainfall used for forcing and measured streamflow for calibration."

Reviewer #3

Comment 1: The primary limitation of this work is the lack of physical interpretation of the outcomes. The hydrological model is treated more as a "black box," failing to provide insights into the hydrological processes that might explain the presented results. I therefore recommend this paper for major revision. In particular, I suggest enhancing the description of the hydrological model, which could assist both the authors and readers in interpreting the results in the discussion section.

Response: Thank you for your suggestion. We recognize that our manuscript lacked a detailed description of the model and a thorough explanation of its outputs. In Section 2.2, we have added a more comprehensive description of the model, including the parameter settings used in this study. Additionally, we selected three representative catchments and have provided a detailed discussion of the differences in simulation results under varying resolution input and streamflow data conditions. We hope this will help clarify the execution, results, and discussion of our experiments.

Comment 2: The methodology section is somewhat difficult to follow. From what I understand, the minimum time step used is one hour, meaning that processes are integrated over time using this interval. This implies that the maximum resolution of the input data is hourly. Once this time step is established, it is possible to produce output at this resolution or at coarser resolutions (e.g., 2 hours, 3 hours, up to 24 hours), but not finer. The discriminating factor is therefore the integration time, after which you may choose to aggregate the output (e.g., streamflow) at a daily time step to compute efficiency metrics. In this context, statements like "all input data except for rainfall were resampled to the hourly resolution" could be misleading.

Response: Your understanding is completely correct. We sincerely apologize for the confusion caused by our wording. We acknowledge that the use of "except for" in our original statement was incorrect; the proper term should have been "besides", and the full sentence should be "All input data besides rainfall were resampled to the hourly resolution".

The purpose of this study is to investigate the value of different data resolutions (including rainfall and streamflow) on hydrological modeling. The original resolution of the rainfall data is slightly finer than 1 hour. We first resampled the rainfall data to other temporal scales between 1 and 24 hours to generate rainfall data at different resolutions. The original resolutions of the other driving data are 24 hours. To eliminate the influence of computation time step, we fixed the model's computation time step at 1 hour. The model was run with a 1-hour computation time step, so if the input data resolution was coarser than 1 hour, it needed to be resampled to a 1-hour resolution before the calculations, and then generate results at a 1-hour resolution. Subsequently, to explore the effect of streamflow data at different resolutions, the 1-hour resolution simulation results were aggregated to other temporal scales, and the model was calibrated based on observed streamflow at the corresponding scales.

In the revised manuscript, we have made appropriate modifications to Section 2.3 on experimental design, including the use of correct wording, emphasizing the purpose, and adjusting the narrative sequence, to enhance reader comprehension of our methodology.

Comment 3: Please provide the equation for the GOUE. Additionally, a table outlining the influencing factors would enhance the readability of the paper.

Response: Thank you for your suggestion. The equation for the GOUE is shown as follows:

$$\text{GOUE} = 1 - \frac{\sum_{i=1}^n (Q_i^{Dh} - Q_i^h)^2}{\sum_{i=1}^n (Q_i^h - \bar{Q}^h)^2}$$

where n is the length of the time series of the streamflow, Q^h is the actual hourly streamflow, \bar{Q}^h is the mean of the Q^h , Q^{Dh} is the hourly streamflow based on daily streamflow assuming a uniform intraday streamflow. We added the GOUE's equation and description to the revised manuscript. Besides, a brief introduction and abbreviations of the influencing factors have been included in Table 1, and we mentioned Table 1 in Section 2.1 of the revised manuscript.

Comment 4: Could the authors discuss other potential factors that might explain the differences in performance between the different basins? For instance, consider a scenario where a particular process is not well-modeled. This process could be highly significant in one basin but negligible in another, leading to varying results. Could this be the case for some of the basins under consideration?

Response: Thank you for your suggestion. As you mentioned, the accuracy of the model's simulation of a particular hydrological process and the importance of that process within the catchment could indeed be one of the potential factors. However, since the meteorological conditions across our study catchments are quite similar, there might not be a significant difference in the importance of a particular hydrological process across different catchments. Therefore, this factor may have little impact on the model's varying performance across the catchments.

Comment 5: The table captions need to be expanded. Specifically, the meaning of bold text and asterisks should be clearly explained in the captions.

Response: Thank you for your suggestion. We followed your suggestion and explained the meaning of the bold text and asterisks in the captions in the revised manuscript.

Comment 6: The second question regarding the time step resolution is not fully addressed. According to the methodology section, the highest resolution used is hourly. With input data of an hourly resolution, one could potentially model processes at this time step.

Response: We apologize for imprecise wording and not clearly explaining the focus of our study. Indeed, if high-resolution data (such as 1-hour resolution) is already available, it can typically provide reliable hourly simulations. However, the purpose of our second question and the design of our second experiment is to determine whether, in scenarios where data availability is limited, relatively coarse resolution data can still offer the same level of reliability in hourly simulations. If so, what is the coarsest resolution that maintains this reliability? We are searching for a resolution threshold: if the data resolution is coarser than this threshold, the reliability of the results decreases; if it is finer, there is no significant improvement in reliability.

We revised the second question to: "What is the coarsest resolution of rainfall and streamflow data required to provide reliable hourly streamflow simulations?" This change, replacing "necessary" with "coarsest," should help reduce confusion and enhance readability.

Comment 7: Line 25-28: References are required to support these statements.

Response: Thank you for your reminding. We have added relevant references to the revised manuscript. The following is the revised content:

To address this limitation, data is often artificially disaggregated from raw time series using mass curves (Blöschl and Sivapalan, 1995) or complex stochastic generators (Creutin and Obled, 1980). However, models based on coarsely resolved or artificially refined data can introduce biases, particularly when forecasting at finer temporal scales, as they may not accurately capture the variability and magnitude of hydrological variables (Younis et al., 2008; Huang et al., 2019).

References:

Blöschl G, Sivapalan M. Scale issues in hydrological modelling: a review[J]. Hydrological processes, 1995, 9(3-4): 251-290.

Creutin J D, Obled C. Modelling spatial and temporal characteristics of rainfall as input to a flood forecasting model[J]. IAHS-AISH Publication (129), 1980: 41-49.

Huang, Y., Bárdossy, A., and Zhang, K.: Sensitivity of hydrological models to temporal and spatial resolutions of rainfall data, *Hydrol. Earth Syst. Sci.*, 23, 2647–2663, <https://doi.org/10.5194/hess-23-2647-2019>, 2019.

Younis J, Anquetin S, Thielen J. The benefit of high-resolution operational weather forecasts for flash flood warning[J]. *Hydrology and Earth System Sciences*, 2008, 12(4): 1039-1051.