

Response to Referee #4

Reviewing of the manuscript 'Towards Improved Euro-Mediterranean Discharge Simulations in Regional Coupled Climate Models: A Comparative Assessment of Hydrologic Performance' by M. Hamitouche et al. submitted Geoscientific Model Development (Manuscript Number: egusphere-2025-2752).

The authors present two updated modules for the hydrological simulation in the coupled earth system (e.g., CaMa-Flood and WRF-Hydro), compared to the default HD models. I am really enjoying reading this modeling paper, since it is clear for its research gap, goal, and limitations. The presentation is clear for the model results. Specifically, two key questions regarding the applicability and improvement of the updated modules to the Euro-Mediterranean regional coupled system have been addressed well with detailed analysis of the multiple statistical formulations and values. I would recommend it for publication after Minor Revisions. The following are my comments and suggestions for the authors to reference:

1. The authors select 11 basins for model performance evaluations, and they are mostly located on the northern border of the Mediterranean Sea. Would the omissions of the basins located on other three sides of the Sea influence the modeling results?

- Thank you for this comment. We acknowledge that most of the selected basins are located along the northern border of the Mediterranean Sea. The inclusion of additional basins from the southern and eastern sides would indeed provide a more spatially complete assessment. However, this was not feasible in the present study due to several constraints already discussed in the manuscript, including limited availability of reliable discharge observations and the strong regulation affecting many rivers in those regions.

Moreover, many of the rivers in these areas have relatively small contributions to the total freshwater flux at the Mediterranean land–ocean interface. Therefore, the selected basins still capture the major hydrological contributions relevant for evaluating model performance at the regional scale.

2. Table 1. Could you double check the definition of beta: whether it is as the authors described or it should be the 'mean simulated minus mean observed and then use this value to be divided by the mean observed'?

- Thank you for your comment.

The definition of β used in Table 1 follows the formulation $\frac{\mu_{sim}}{\mu_{obs}}$, i.e., the ratio between the mean simulated discharge and the mean observed discharge. Therefore, β represents the multiplicative bias factor and is calculated as the mean simulated divided by the mean observed, as stated in the manuscript.

We have carefully checked the definition and confirm that it is correctly reported in the paper.

3. Table 2. It should be noted that if the largest basin Danube being excluded, the averaged lag values for CaMa-Flood model even better than the WRF-Hydro (Default parameters) (e.g., 5.0 days versus 5.3 days). This conclusion is not consistent with the overall conclusion which the

Danube is included. The reviewer consider that this information could be indicated so that the readers could have a comprehensive idea of the two-model comparison.

- Thank you for this insightful comment. We agree that the large lag simulated by CaMa-Flood in the Danube basin strongly influences the averaged lag values. As suggested, we have clarified this point in the manuscript (Lines 364-366) by indicating that when the Danube basin is excluded, the average lag becomes very similar between the two models (5.0 days for CaMa-Flood vs. 5.3 days for WRF-Hydro).

4. The KGE value is popular and important as the metric skills. The reviewer considers that it is better to give the definition of the excellent/good, normal, poor etc. categories correspondence to the values.

- Thank you for this suggestion. We agree that providing performance categories for the KGE metric helps readers better interpret the results. Accordingly, we have added a clarification in the manuscript (Lines 295–300) in Section 2.3 describing the interpretation ranges of KGE values.

Specifically, we now indicate that KGE values ≤ -0.41 are considered unacceptable, $-0.41 < KGE \leq 0.00$ very poor, $0.00 < KGE \leq 0.30$ poor, $0.30 < KGE \leq 0.65$ intermediate, and $0.65 < KGE \leq 1.00$ good, following the benchmark interpretation proposed by Knoben et al. (2019) and the classification adapted from Sanchez Lozano et al. (2025).

5. Line 436: '..... as moderate differences in extremes do not strongly affect long-term water balance'. Does this statement have any supporting literatures and it is a popular fact?

- Thank you for this comment. The statement refers to the fundamental hydrological water balance assumption, where total inflows (e.g., precipitation) must equal total outflows (e.g., runoff and evapotranspiration) plus changes in water storage. When considering sufficiently long time periods (typically longer than one year), changes in water storage are commonly assumed to be negligible ($\Delta S \approx 0$). Under this assumption, moderate differences in the representation of extremes generally have limited influence on the long-term water balance.

To support this statement, we have added appropriate references (Hamitouche et al., 2025; Oda et al., 2024) in the manuscript.

6. Could Fig.4 and 6 be merged? Also, for Fig. 3 and 7, Table 2 and 3, Fig. 5 and 8?

- Thank you for this suggestion. Following the reviewer's recommendation, we have merged Figures 4 and 6, Figures 5 and 8, as well as Tables 2 and 3 in the revised manuscript to improve conciseness.

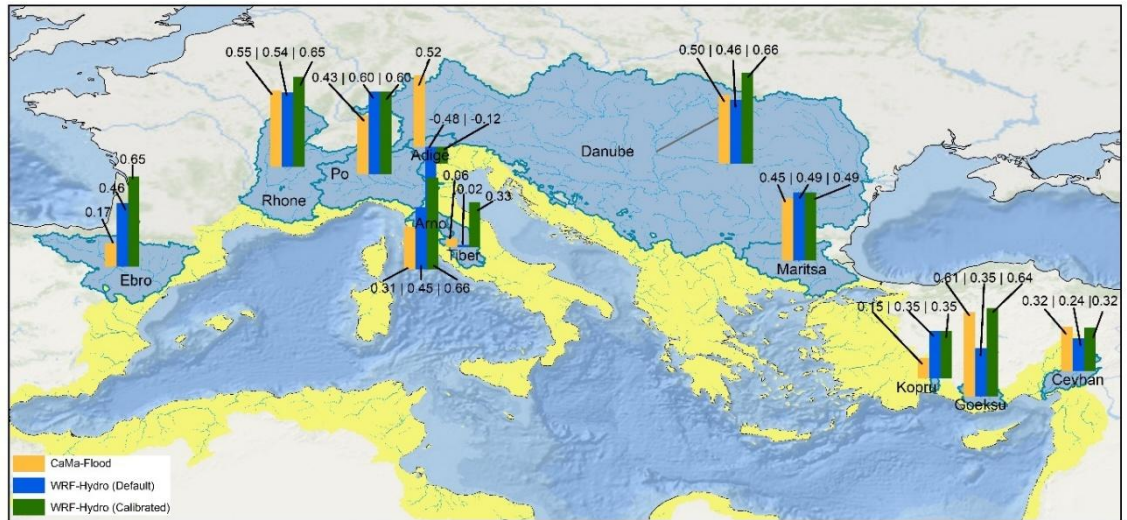


Figure 4: KGE values of each model experiment (CaMa-Flood; WRF-Hydro (Default); WRF-Hydro (Calibrated)) across the studied basins.

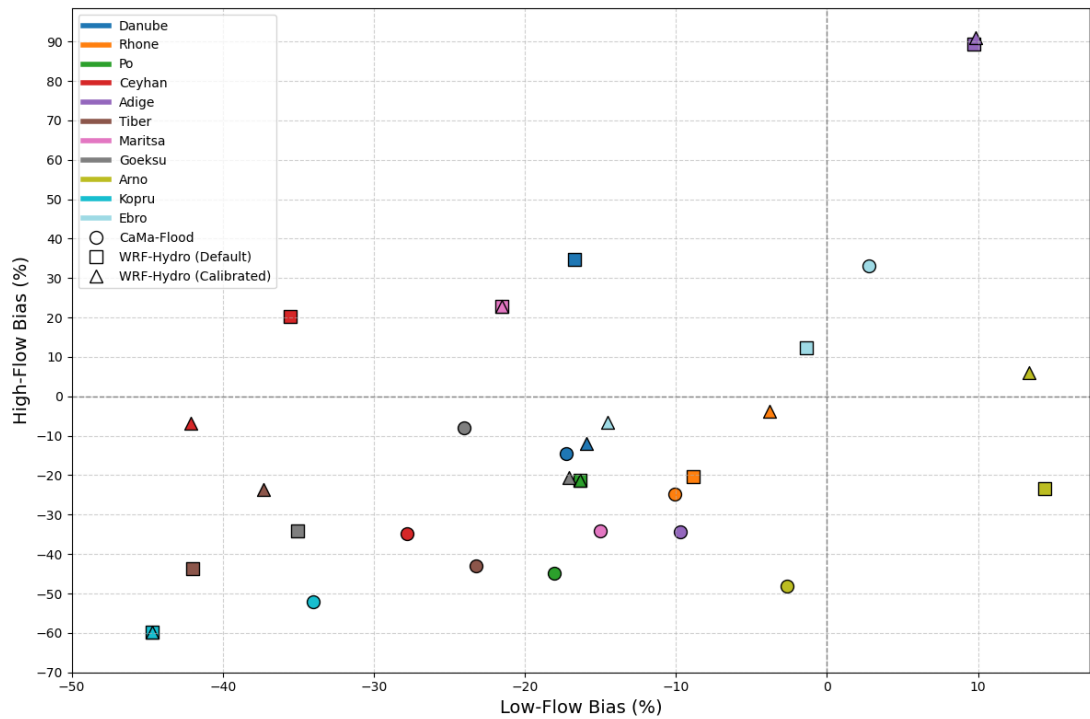


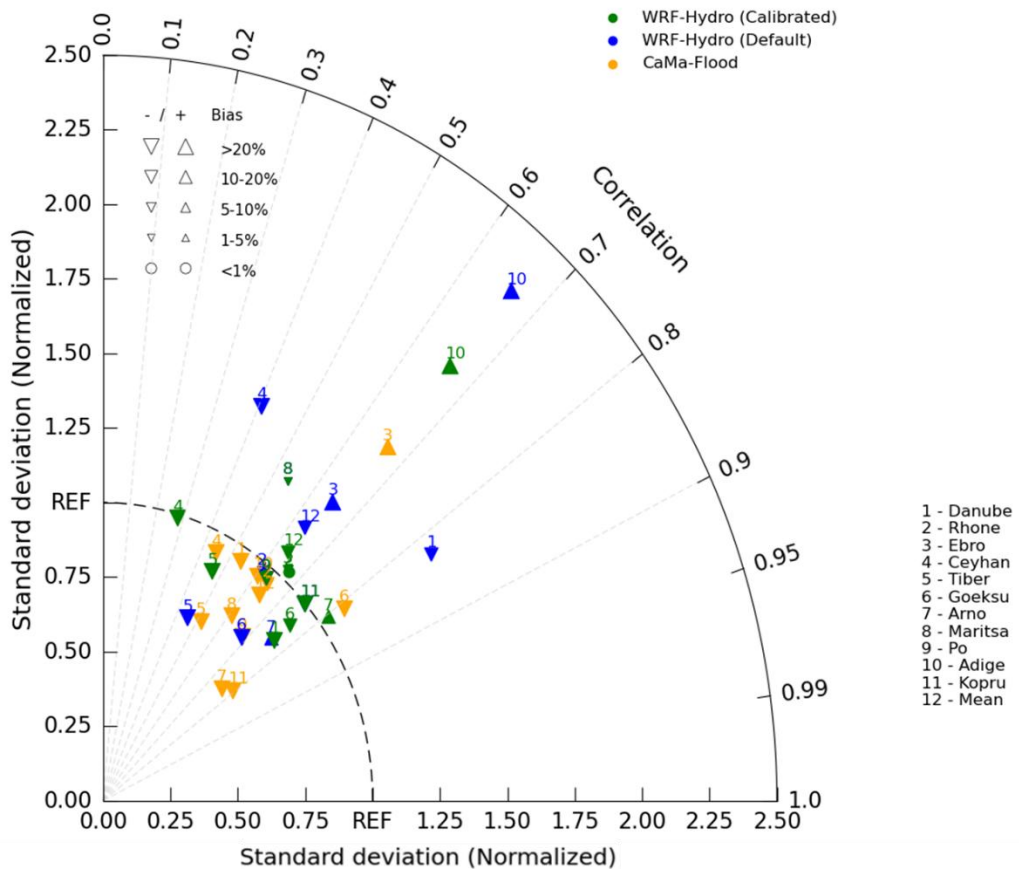
Figure 5: Scatterplot of low-flow bias vs. high-flow bias in each model experiment (CaMa-Flood; WRF-Hydro (Default); WRF-Hydro (Calibrated)) and river basin.

Table 1: Summary of time lag values for each model experiment (CaMa-Flood; WRF-Hydro (Default); WRF-Hydro (Calibrated)) and each basin.

Basin	CaMa-Flood	WRF-Hydro (Default parameters)	WRF-Hydro (Calibrated)
Danube	43	0	3

Rhone	5	6	1
Po	6	2	
Ceyhan	3	3	0
Adige	4	3	0
Tiber	6	9	0
Maritsa	8	7	
Goeksu	1	6	3
Arno	6	6	0
Kopru	1	4	
Ebro	10	7	0

However, we decided to keep Figures 3 and 7 separated. When combined, the figure becomes visually crowded and some model values become more difficult to distinguish clearly. For this reason, we believe that keeping these figures separate improves readability. An example of the merged version is shown below for reference.



7. The results present very well, while more deep discussion could make this paper stronger in '3 Results and discussion' Section. For example, compare with resent and current similar researches and literatures.

- *Thank you for this helpful suggestion. To strengthen the discussion, we have expanded Section 3 (Results and Discussion) to include comparisons with recent studies that evaluated hydrological model performance using WRF-Hydro. These studies report comparable overall behaviour, with WRF-Hydro generally capturing flow timing and variability reasonably well but sometimes overestimating peak flows, which is consistent with the patterns observed in our results.*

Regarding CaMa-Flood, very few studies have evaluated its routing performance over Euro-Mediterranean basins. Existing regional applications have mainly used the model for large-scale water cycle analyses rather than detailed basin-scale discharge performance assessments. We have therefore clarified this point in the manuscript and discussed how our findings contribute new insights into the behaviour of CaMa-Flood in this regional context.

Relevant references have been added in the revised manuscript to support this comparison with the existing literature:

“Although direct comparisons with previous studies are limited due to differences in model configuration, spatial resolution, forcing data, and basin characteristics, the overall behaviour observed here is broadly consistent with the literature. Applications of WRF-Hydro in various hydrological contexts have generally shown that the model can reproduce streamflow dynamics with reasonable skill, particularly after calibration. For example, coupled and uncoupled implementations of WRF-Hydro have reported efficiency metrics ranging from moderate to high performance (e.g., NSE or KGE values around 0.7–0.8 in calibrated applications), demonstrating the model’s capability to capture the timing and magnitude of discharge in different basins (Ndiaye et al., 2026; Senatore et al., 2015; Wang et al., 2020). However, several studies also highlight persistent challenges in reproducing peak flows or extremes, often linked to precipitation forcing errors or parameter uncertainties (Naabil et al., 2017). The behaviour identified in the present study—where WRF-Hydro better captures variability but may occasionally overestimate peaks—therefore aligns with the general characteristics reported in previous evaluations.

In contrast, very few studies have assessed the performance of CaMa-Flood in the Euro-Mediterranean context. Existing applications in this region have mainly used the model for large-scale water-cycle analyses or river discharge estimates rather than detailed basin-scale performance evaluation (e.g., closing the regional water balance), limiting opportunities for direct benchmarking of routing performance. Consequently, the systematic underestimation of variability and extremes observed here provides new insight into the behaviour of CaMa-Flood when driven by a regional Earth system model over Mediterranean basins.”

8. Would it be helpful by adding one table showing the key parameters settings between the default and calibrated WRF-Hydro models?

- *Thank you for this suggestion. The key parameter settings for the default and calibrated WRF-Hydro configurations are already provided in the Supplement (Table S7).*