Response to Anonymous Referee #1

The authors offer a regionally-focused comparison of two hydrological routing models, CaMa-Flood and WRF-Hydro, using default parameters, as well as a calibrated WRF-Hydro. The paper is a nice contribution to the Mediterranean region hydro modeling scientific literature. I suggest Major Revision, with three Major comments and several Minor comments.

Major comments:

1) **Compare results with a baseline**. To show improvement in river discharge simulations, it would be helpful to compare the results with an existing baseline/benchmark. For instance, in the introduction, part of the motivation was that the HD model was underestimating discharge in the Med. Similarly, CaMa and WRF-Hydro also tended to underestimate discharge. Is there any way to see if either model actually offered an improvement? Even if it is qualitative? If not, this point should still be mentioned as a limitation of the work.

- ***We thank the reviewer for this valuable suggestion. To address it, we added a dedicated discussion of the HD model (which is the default river-routing component in the current ENEA-REG configuration) as a baseline benchmark in Section 3 ("Results and Discussion"). Specifically, before comparing the performances of CaMa-Flood and WRF-Hydro, we now present both a qualitative and quantitative evaluation of HD simulations against observations. This includes discharge time series (as Fig. S4 in the supplement) to visually contrast HD results with observations, CaMa-Flood, and WRF-Hydro, and statistical metrics averaged across all Mediterranean basins, with basin-scale details provided in the Supplement (as Tables S4 and S5).***

3 Results and discussion:

"Before comparing the performances of CaMa-Flood and WRF-Hydro, we first present their simulated discharge time series, alongside observations and the discharge simulated by the HD model at both 0.5-degree and 5-minute spatial resolutions (Fig. S4). The higher-resolution HD configuration is intended to improve the representation of the drainage network and catchment area, thereby enhancing discharge estimates. However, both HD versions show a clear underestimation of freshwater inflow to the Mediterranean Sea and fail to reproduce the observed temporal patterns.

Quantitatively, the HD model exhibits very poor performance across Mediterranean basins. At 0.5-degree, the overall average KGE is –0.13, with a mean bias of –37.4%, a Spearman's correlation of 0.36, and a relative SD of only 0.23. At 5-minute, the performance improves slightly (KGE = 0.00; %Bias = –16.6%; r_s = 0.48), but variability remains strongly damped (rSD = 0.25). The most critical shortcoming is the systematic underestimation of extremes: high-flow biases reach –70.3% (0.5-degree) and –65.7% (5-minute), while low-flow biases are also negative (–25.3% and –9.2%, respectively). Model performances at the basin scale for both HD versions (0.5-degree and 5-minute) are provided in Tables S5 and S6 of the Supplement, respectively, offering basin-specific details beyond the aggregated statistics presented here.

These results suggest that the underestimation of freshwater inflow by the HD model is not primarily due to the long-term mean bias—which is moderate in the 5-minute configuration—but rather to its inability to capture discharge variability and high-flow peaks.

In contrast, visually WRF-Hydro and CaMa-Flood reproduce observed temporal dynamics much more closely, including variability and extremes, and therefore clearly outperform the HD

model. Both models show strong potential for capturing observed patterns when using runoff from the Noah-MP land surface model—the default LSM in many Euro-Mediterranean regional coupled and Earth system models—which has already been validated against ERA5-Land runoff data (Hamitouche et al., 2025). This indicates that CaMa-Flood and WRF-Hydro are suitable alternatives to replace the HD model in this modelling framework, despite some visual biases in low- and high-flow reproduction. The following section provides a detailed evaluation of their performances."

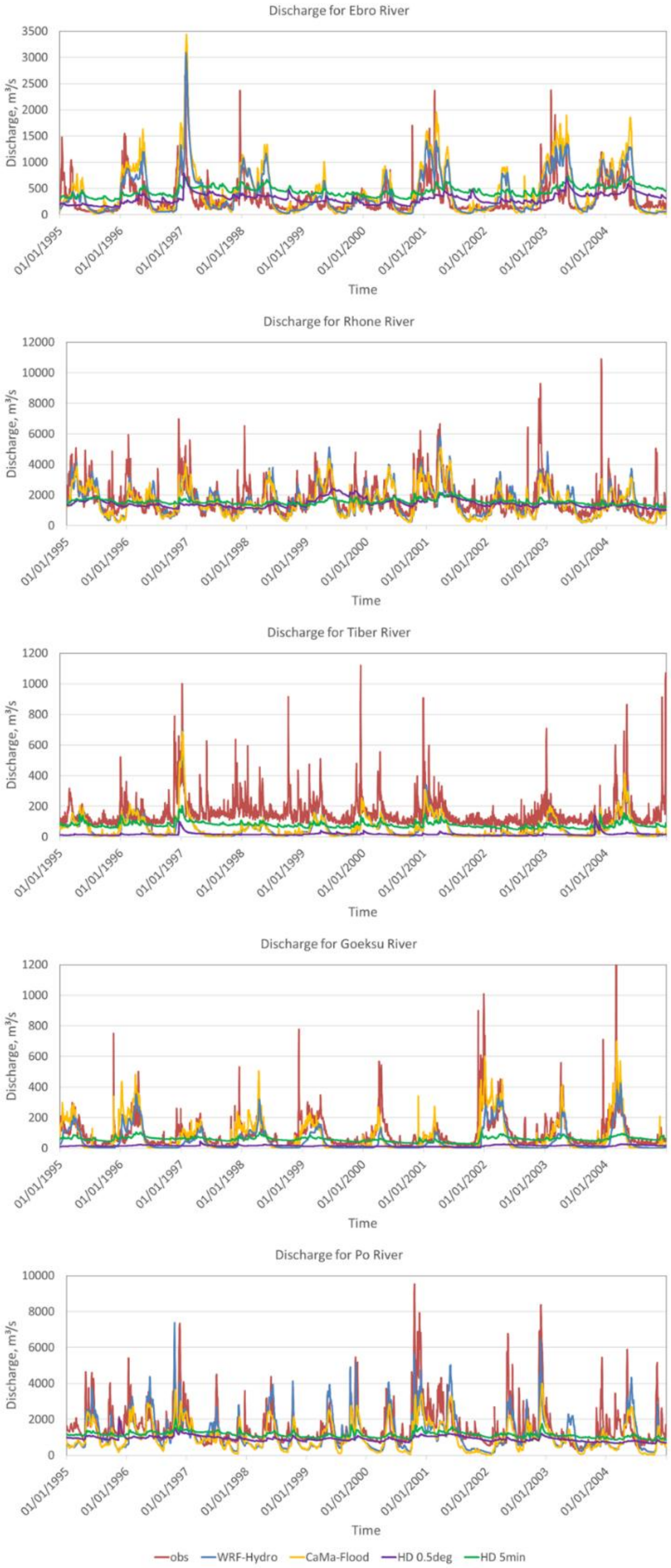
Discharge for Ebro River
Discharge, m³/s
Time
Discharge for Rhone River
Discharge, m³/s
Time
Discharge for Tiber River
Discharge, m³/s
Time
Discharge for Goeksu River
Discharge, m³/s
Time
Discharge for Po River
Discharge, m³/s
Time
obs
WRF-Hydro
CaMa-Flood
HD 0.5deg
HD 5min

Figure S4: Observed and simulated daily discharge for the Ebro, Rhone, Tiber, Goeksu and Po rivers for common 10 years from 1995 to 2004. Simulations include ENEA-REG–driven WRF-Hydro and CaMa-Flood, as well as the HD model at 0.5° and 5-minute spatial resolutions, evaluated near the corresponding gauge stations.

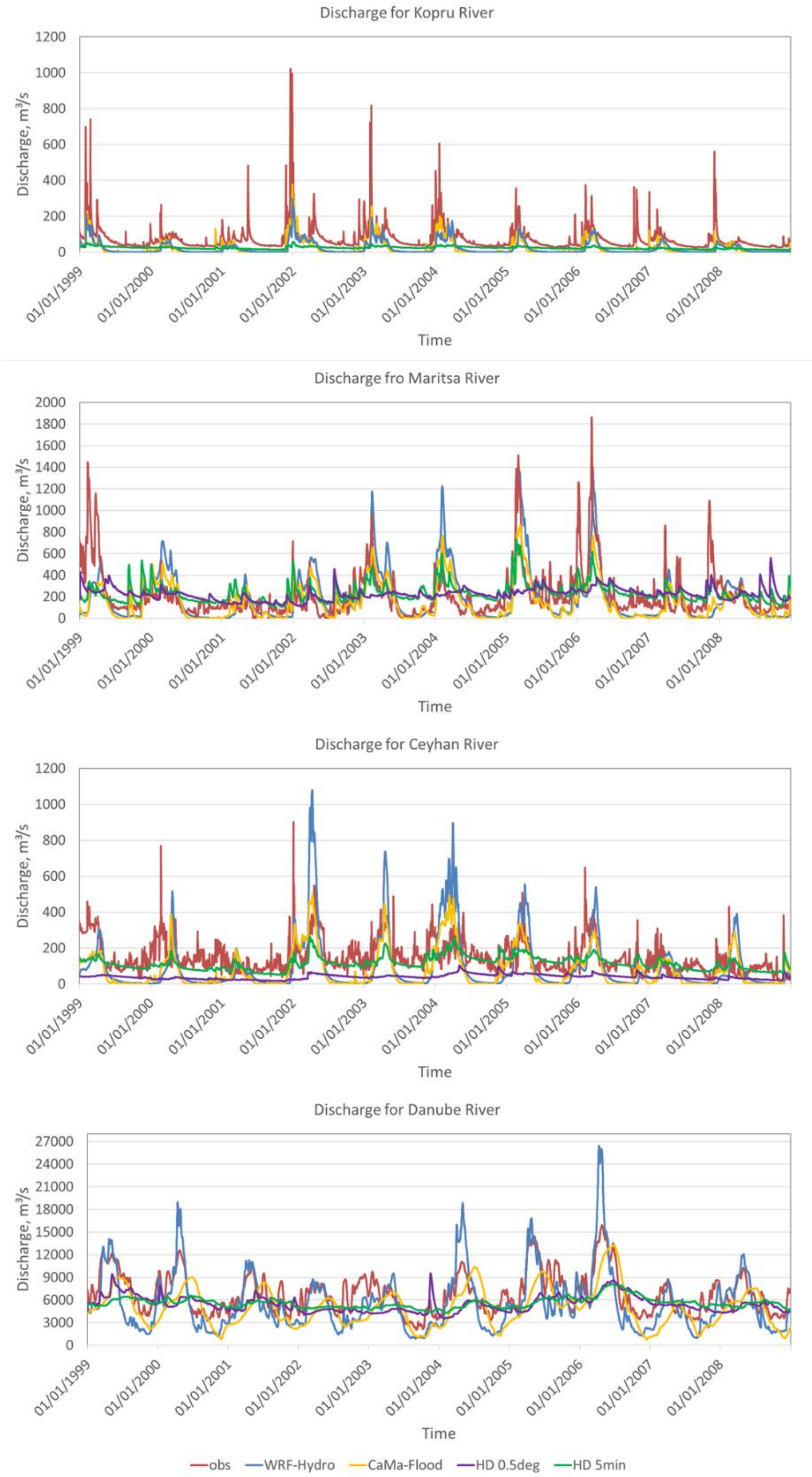

Figure S4 (continuity): Observed and simulated daily discharge for the Kopru, Maritsa, Ceyhan and Danube rivers for common 10 years from 1999 to 2008. Simulations include ENEA-REG–driven WRF-Hydro and CaMa-Flood, as well as the HD model at 0.5° and 5-minute spatial resolutions, evaluated near the corresponding gauge stations.

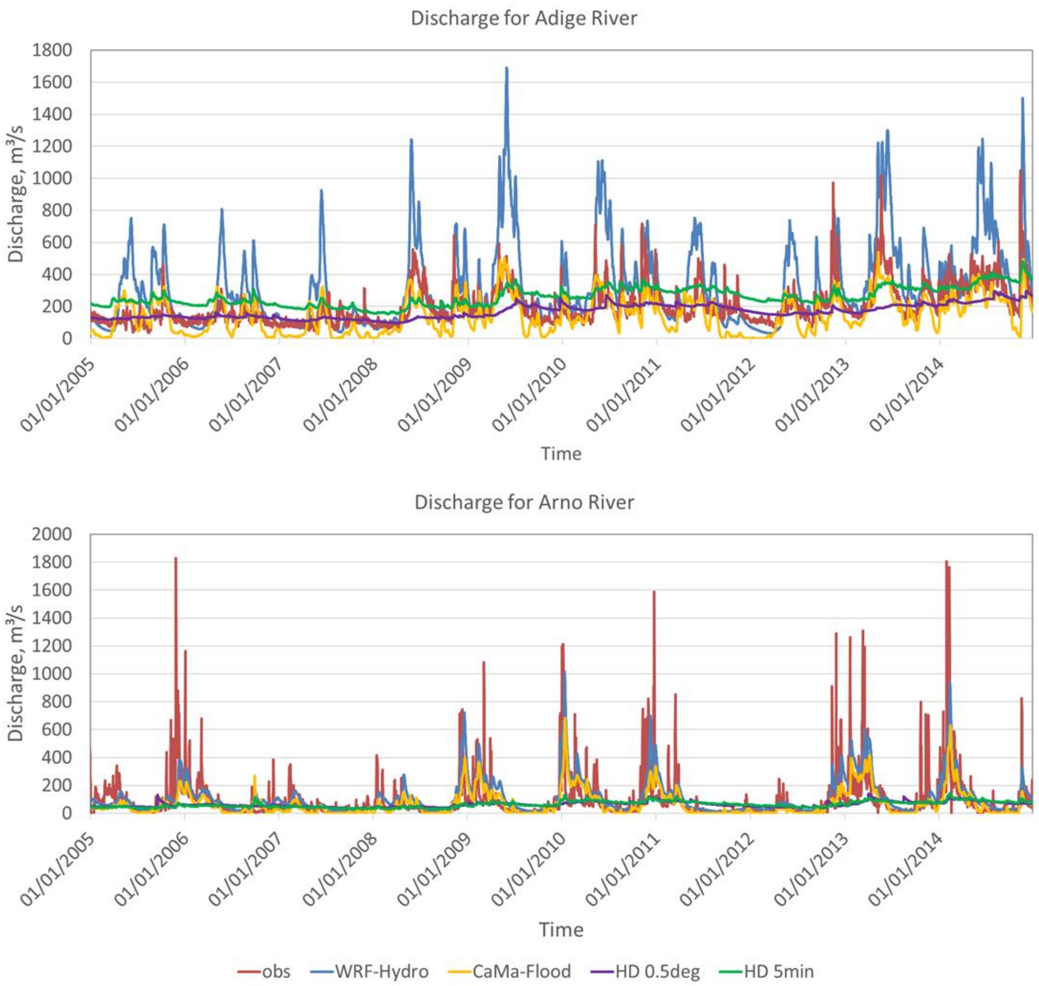

Figure S4 (continuity): Observed and simulated daily discharge for the Adige and Arno rivers for common 10 years from 2005 to 2014. Simulations include ENEA-REG–driven WRF-Hydro and CaMa-Flood, as well as the HD model at 0.5° and 5-minute spatial resolutions, evaluated near the corresponding gauge stations.

Table S5: Summary of the performances (KGE, bias, correlation, relative standard deviation, low-flow and high-flow biases) of HD model at 0.5-degree spatial resolution for each river basin, evaluated over the entire 1990-2014 period. No values are reported for the Kopru basin, as its small size relative to the coarse 0.5-degree resolution prevented extraction of a representative discharge time series.

Basin	KGE	Percent bias	Spearman's rho	Relative SD	High-flow Percent Pbias	Low-flow Percent Pbias
Danube	**0.15**	**-21.88**	**0.35**	**0.42**	**-40.62**	**-12.44**
Rhone	**-0.05**	**-20.92**	**0.31**	**0.28**	**-61.07**	**-18.73**
Po	**-0.08**	**-42.61**	**0.52**	**0.18**	**-77.32**	**-42.33**
Ceyhan	**-0.36**	**-75.77**	**0.30**	**0.11**	**-88.31**	**-63.25**
Adige	**0.11**	**-14.57**	**0.47**	**0.33**	**-57.16**	**-1.36**

Tiber	-0.47	-89.30	0.40	0.10	-89.02	-67.41
Maritsa	-0.01	1.56	0.32	0.27	-65.12	6.91
Goeksu	-0.49	-83.36	0.26	0.05	-94.07	-66.58
Arno	-0.18	-29.77	0.32	0.13	-84.83	8.13
Kopru			/			
Ebro	0.10	2.86	0.33	0.45	-45.08	4.42

Table S6 Summary of the performances (KGE, bias, correlation, relative standard deviation, low-flow and high-flow biases) of HD model at 5 minutes spatial resolution for each river basin, evaluated over the entire 1990-2014 period.

Basin	KGE	Percent bias	Spearman's rho	Relative SD	High-flow Percent Pbias	Low-flow Percent Pbias
Danube	-0.02	-19.02	0.21	0.30	-45.14	-8.82
Rhone	-0.05	-15.30	0.31	0.22	-64.42	-6.55
Po	-0.01	-29.79	0.53	0.21	-74.04	-21.67
Ceyhan	0.12	-27.36	0.50	0.32	-64.56	-7.97
Adige	0.18	38.65	0.55	0.44	-37.52	16.82
Tiber	-0.07	-52.18	0.60	0.18	-78.90	-59.56
Maritsa	0.31	6.41	0.64	0.40	-50.96	7.33
Goeksu	-0.05	-36.57	0.58	0.17	-80.35	-6.63
Arno	-0.13	-28.15	0.34	0.13	-84.43	9.83
Kopru	-0.22	-68.48	0.64	0.10	-89.17	-61.27
Ebro	-0.05	49.63	0.42	0.33	-53.01	37.40

2) **Integrate the calibration results**. As written and presented, the calibrated WRF-Hydro results seem like a separate add-on, rather than an integrated component. Some integration occurred starting on line 466, but this was hard to follow.

- ***We appreciate the reviewer's observation. Our intention was to present the calibration of WRF-Hydro as a complementary analysis rather than as part of the main model comparison. This decision was motivated by several factors:***
  1. ***Unlike WRF-Hydro, CaMa-Flood and Noah-MP do not currently support basin-specific or spatially variable parameter calibration, and Noah-MP (within ENEA-REG) is not two-way coupled with CaMa-Flood. To ensure fairness in the core comparison between the two routing models, we restricted the main evaluation to non-calibrated configurations.***

2. ***WRF-Hydro calibration was facilitated thanks to the availability of the PyWRFHydroCalib package, whereas no equivalent tool exists for CaMa-Flood in our framework.***
3. ***Calibration is computationally expensive, and we considered it more appropriate to present the results as an "add-on" analysis, illustrating the potential benefits of calibration where resources allow, rather than as a standard step in the intercomparison.***

For these reasons, we structured the calibration results as a separate subsection, highlighting their added value while keeping the baseline comparison consistent and unbiased. We have revised the section headings to clarify this rationale and improve the readability of the calibration analysis.

3) **Improve clarity, organization, and flow.** The authors include a lot of good content, but sometimes it was hard to follow and assumed prior familiarity. Specific suggestions for this are offered in the Minor Comments listed line-by-line, below.

- ***We thank the reviewer for this observation. We carefully revised the manuscript by addressing the suggestions provided in the minor comments, which helped us improve the clarity, organization, and flow of the paper.***

Minor comments:

1) Line 25. Suggest: "Not only does it provide essential freshwater input…"

- ***We thank the reviewer for the suggestion. The sentence has been revised accordingly to: "Not only does it provide essential freshwater input…"***

2) Line 30: Suggest tightening up this paragraph to better link with previous paragraph, which was only about the importance of the regional discharges in the Med Sea. Suggest revising this first sentence (also not sure if the Nearing reference fits, since that ref is more about AI-forecasting). Maybe: "Timely and accurate river discharge estimates into the Mediterranean are critical for managing water resources and risks in the region." Then second sentence could merge with the 3rd to be, "In a broader context, understanding the interplay between hydrological processes and regional climate is essential, and underscores the need to study the coupling of …."

- ***We thank the reviewer for the helpful suggestion. The paragraph has been revised to improve clarity and linkage, and now reads:***
  "Timely and accurate river discharge estimates into the Mediterranean are critical for managing water resources and related risks in the region (Cisterna-García et al., 2025). In a broader context, understanding the interplay between regional climate change and hydrological processes is essential, and underscores the need to study the coupling of Earth system components, including atmospheric, hydrological, and oceanic processes.".

3) Line 56. Suggest replacing "Nowadays". Maybe something about Recent advances?

- ***We thank the reviewer for the suggestion. The sentence has been revised to:***
"Recent advances have led to the development of several complex coupled models, aimed at achieving fully integrated hydrological predictions for the Mediterranean region."

4) Line 60-64. This has important information but is hard to follow for someone not very familiar with these models; suggest revising for clarity. What is HD? How do the different horizonal resolutions factor in? What is HydroPy LSM?

- ***We thank the reviewer for pointing this out. To improve clarity, we first explicitly define the models: the HD model refers to the Hydrological Discharge model (Hagemann and Dümenil, 1997), which is a river routing model, while HydroPy is a Land Surface Model (LSM) (Stacke and Hagemann, 2021) that simulates runoff.***
The revised text now reads:
"This underestimation seems to be with the Hydrological Discharge (HD) model (Hagemann and Dümenil, 1997), a river routing model used to simulate the river discharge with different horizontal resolutions (e.g. 5 minutes in MESMAR and 0.5 degrees in ENEA-REG coupled models). The HD model uses a pre-parametrization based on a linear reservoir routing concept with pre-defined reservoir numbers and temporal constants tailored to runoff inputs from the HydroPy Land Surface Model (LSM) (Stacke and Hagemann, 2021), which neglects the energy budget and overestimates runoff."
We also clarify later in Section 3 ("Results and discussion") that the higher-resolution HD configuration (5-minute) is intended to improve the representation of the drainage network and catchment area, thereby enhancing discharge estimates (see reply to major comment 1).

5) Line 75. About this sentence -> "This highlights the importance of conducting detailed sensitivity analyses on standalone routing models, before adding them into coupled climate or Earth system models." <- I think this is too general, and what is the sensitivity analysis here? Seems more like a comparative evaluation. Also, can you link this to a specific regional model or modelling experiment, such as the Med-CORDEX experiment, rather than the general coupled ESMs? Or later when you say "Euro-Mediterranean regional coupled models"?

- ***We thank the reviewer for pointing this out. We agree that the original statement was too general and could be misleading. To address this, we revised the sentence to focus on the comparative evaluation carried out and explicitly linked it to the Med-CORDEX initiative. The revised text now reads:***
"This highlights the importance of thoroughly evaluating standalone routing models for their performance and limitations, before integrating them into regional coupled climate or Earth system models—such as the ENEA-REG atmosphere-land-river-ocean (ALRO) coupled model, developed within the framework of the Mediterranean CORDEX (Med-CORDEX) initiative."

6) Line 84. This is the first time you mention Med-CORDEX. Even though you define CORDEX earlier, you assume that the reader is familiar with this particular experiment. Suggest adding more context for the reader. Maybe also point to Figure 1 here.

- ***We thank the reviewer for this suggestion. In response, we added context about the Med-CORDEX initiative earlier in the manuscript (immediately after the revised line 75), explaining its aims and geographic domain, with the following text:***
  "Med-CORDEX aims to advance fully coupled regional climate simulations over the Euro-Mediterranean domain, which encompasses the Mediterranean and Black Seas and their contributing catchments (excluding the Nile), by improving the representation of key Earth system components, including atmospheric processes, land surface, hydrology and ocean dynamics (Ruti et al., 2016)."
  Additionally, we now explicitly point to Figure 1 at the mention of the "Med-CORDEX domain" (line 84), so that the reader can better understand the spatial context.

7) Line 91-93. The calibration seems like an add-on as written. Can you better integrate this into your experimental design? Why don't you do calibration experiments on CaMa-Flood?

- ***We thank the reviewer for raising this point. This concern was also addressed in our response to major comment 2. In short, calibration was presented as a complementary "add-on" analysis because (i) CaMa-Flood and Noah-MP do not support basin-specific calibration in our framework, (ii) calibration of WRF-Hydro was facilitated thanks to the PyWRFHydroCalib package, and (iii) calibration is computationally expensive. For these reasons, we kept the baseline intercomparison restricted to non-calibrated configurations to ensure fairness, while presenting WRF-Hydro calibration separately to illustrate its added value.***

8) Line 95-96. Going forward, to show improvement, it's best to compare to an existing baseline (such as the Med-CORDEX simulations, or the underestimating HD model you talk about earlier).

- ***We thank the reviewer for this suggestion. Following the same concern raised in the major comment on including a baseline, we have now integrated the HD model as a benchmark in Section 3 ("Results and Discussion"). Both qualitative (time series in Fig. S4) and quantitative (performance metrics in Tables S5 and S6) comparisons with HD are presented, highlighting its underestimation of discharge and providing a clear baseline against which CaMa-Flood and WRF-Hydro improvements can be assessed.***

9) Line 116: This paragraph doesn't fit as-is under the 2.1 heading (Study area and river discharge observations). If it does stay, then this is the first time the ENEA-REG model is mentioned, and the reader needs it to be defined, and more information about what it is. Looks like you don't say it's the atmosphere-land-ocean model until line 135 – should this information be moved to that section? Line 119: You mention the Med-CORDEX protocol – should some of the protocol info be mentioned in the introduction to help the reader with background?

- ***We thank the reviewer for this observation. We chose to keep the paragraph under Section 2.1. To improve clarity, we have defined ENEA-REG at its first mention in the introduction as a coupled model together with MESMAR. Later in the introduction, before providing additional context on the Med-CORDEX initiative, we clarified that ENEA-REG is an atmosphere–land–river–ocean (ALRO) coupled model developed within the Med-CORDEX framework (see the reply to minor comment 5).***
  As suggested, information on the Med-CORDEX protocol has also been provided immediately after the context on the Med-CORDEX initiative:
  "The Med-CORDEX Phase 3 protocol outlines common requirements for domain extent, spatial resolution, and coupling strategies—mandating, for example, a minimum resolution of 12 km for the atmosphere/land and 10 km for the ocean, and requiring river-to-ocean coupling. Further details on the protocol are available at https://zenodo.org/records/11659642.".

10) Line 188: Why is the CaMa model not calibrated and compared during this study? Do you only calibrate wrf-hydro since the PyWrfHydroCalib package is available?

- ***We thank the reviewer for this question. As noted in our reply to Major Comment 2 and Minor Comment 7, unlike WRF-Hydro, CaMa-Flood and Noah-MP do not currently support basin-specific or spatially variable parameter calibration, and Noah-MP (within ENEA-REG) is not two-way coupled with CaMa-Flood. In addition, calibration of WRF-Hydro was facilitated by the availability of the PyWRFHydroCalib package, whereas no equivalent tool exists for CaMa-Flood. For these reasons, only WRF-Hydro was calibrated and compared in this study.***

11) Figure 2, Figure 3, Table 2 etc: Is there a reason to not present all the results together? So include the three (i) default CaMa, (ii) default WRF-Hydro, and (iii) calib WRF-Hydro?

- ***We thank the reviewer for this suggestion. They are not presented together for the fact that the calibration was considered as a separate section for the reasons mentioned previously (see reply to major comment 2, minor comment 7, and minor comment 10), so that the related results are presented in that section. Moreover, when integrating all the results together, some figures—specifically the Taylor diagram—became overloaded, making it difficult to read the results and distinguish the points.***

12) Line 310. You mentioned earlier that the HD model also tends to underestimate discharge. Is there a way to compare how much HD underestimates vs these models? It can help if there is a prior baseline to compare to, even if only qualitatively.

- ***We agree with the reviewer's point. As part of the added baseline analysis (see reply to major comment 1), we now explicitly quantify the HD model underestimation relative to observations, using both 0.5-degree and 5-minute resolutions. These results (Section 3; Fig. S4; Tables S5 and S6) demonstrate that the underestimation is most pronounced in terms of variability and high-flow reproduction, and they serve as a direct benchmark against which WRF-Hydro and CaMa-Flood performance improvements can be evaluated.***

13) Line 434. This seems better to either go earlier (after Figure 5 and the KGE results) or in a discussion section, otherwise it seems out of place and hard to follow.

- ***We thank the reviewer for the helpful suggestion. The line and related sentences have been moved to follow the discussion of Figure 5 and the KGE results.***

14) Line 464. This is a new topic and I recommend adding a new subsection here, and some introductory sentences to guide the reader on what was done and why. Also, I'm not following what is being shown in Table S5. Line 468 – I suggest reminding the reader that the ENEA-REG model is downscaled with WRF. The Table S5 caption needs to also remind the reader that WRF is downscaling the ENEA-REG. I had trouble crosswalking what was in the text and in the supplemental Table S5, suggest revising. Is there a better/clearer way to show the results in Table S5, and/or should it be brought into the main manuscript rather than the Supplemental?

- ***We thank the reviewer for this valuable suggestion. Following the advice, we introduced a new subsection titled "3.2.2 Effects of calibration on runoff generation and partitioning", with an introductory paragraph to guide the reader on what was done and why:***
  "To better understand the source of performance improvements following calibration, we investigated how the calibrated parameters influenced key hydrological processes—particularly runoff generation and the surface–subsurface flow partitioning. This analysis helps explain the hydrological changes driving model behaviour across the Mediterranean basins. Specifically, we examined simulated total runoff, as it serves as the primary input for CaMa-Flood, and the ratio of subsurface to total runoff. As a reminder, runoff is simulated by the Noah-MP land surface model, which is integrated within WRF-Hydro and embedded in WRF—used to dynamically downscale ERA5 data within the ENEA-REG coupled model."
- ***To better show the results previously in Table S5, we replaced the table with a new figure brought into the main manuscript. The figure consists of two subplots using grouped bar charts: subplot (a) represents average total runoff and subplot (b) represents the subsurface-to-total runoff ratio, with each group of bars corresponding to a river basin.***

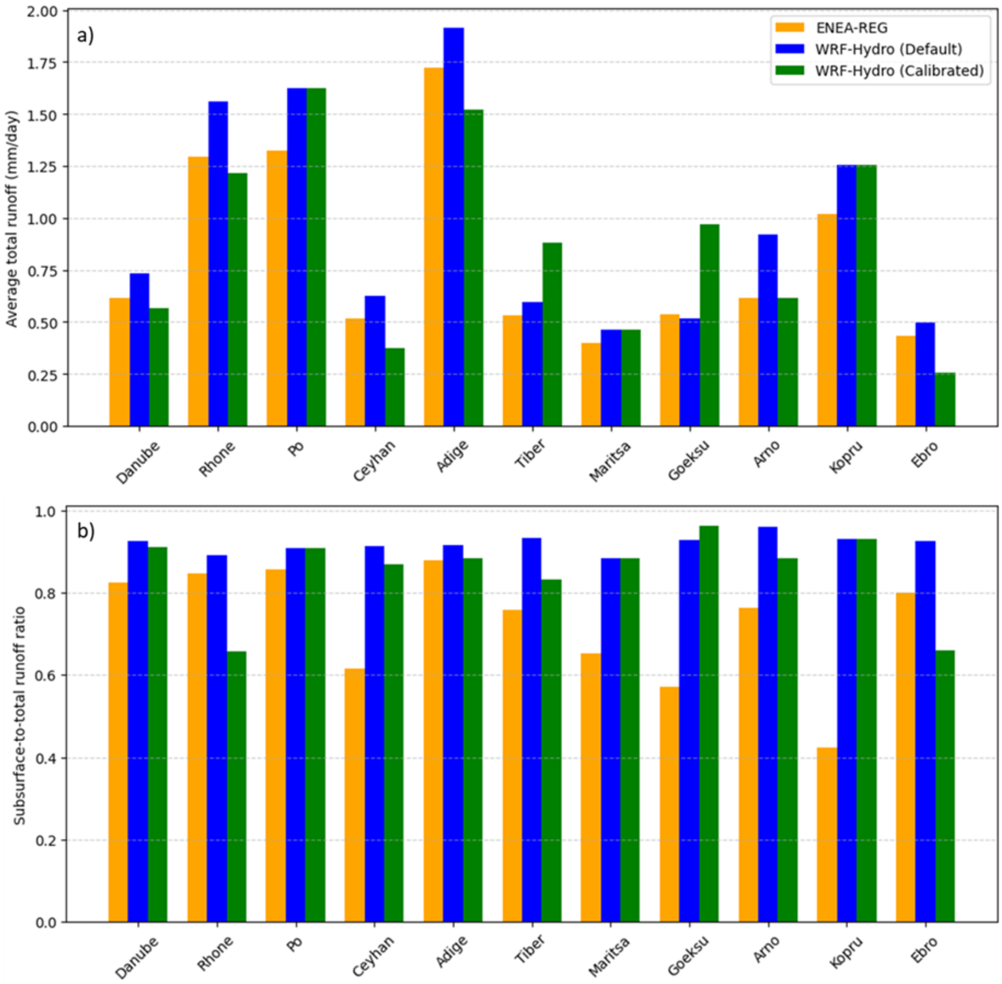

Figure 9: Comparison of a) average daily total runoff and b) the ratio of subsurface to total runoff across each river basin as simulated by ENEA-REG (orange), WRF-Hydro with default parameters (blue), and WRF-Hydro with calibrated parameters (green).

- ***Actually, the ENEA-REG model is not downscaled with WRF, but rather WRF is the atmospheric component of ENEA-REG, used to dynamically downscale ERA5 meteorological data. For this reason, and to avoid confusion, the new figure caption does not mention "WRF" but instead refers to ENEA-REG. At the same time, the introductory paragraph (above) includes a reminder explaining that WRF is used to downscale ERA5 data within ENEA-REG.***

15) Line 493. I don't see a Fig S7.

- ***We thank the reviewer for the observation. Figure S7 is included in the Supplement, immediately following Figure S6.***

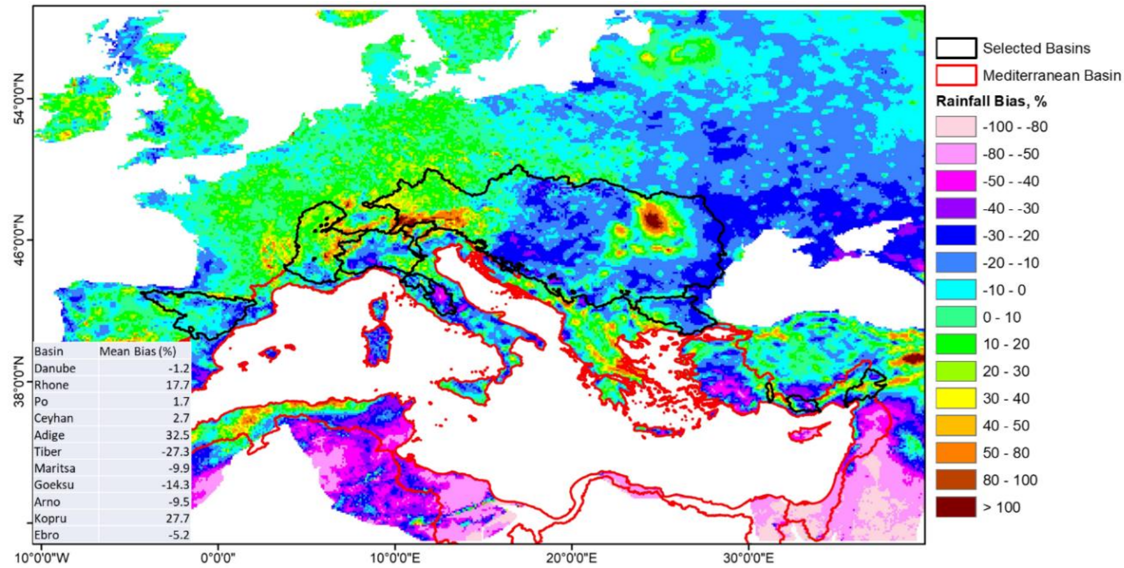

Figure S7: Distribution of the mean daily rainfall percent bias (100 x (WRF-EOBS)/EOBS)

16) Line 495. Would be good to recognize that including a baseline for Euro-Med models would be helpful to see if there was an improvement.

- ***We thank the reviewer for this helpful suggestion. The revised manuscript now explicitly incorporates the HD model as a baseline benchmark for Euro-Mediterranean simulations (Section 3; Fig. S4; Tables S5 and S6). This addition addresses the reviewer's point by providing a clear reference framework, making it possible to show that both CaMa-Flood and WRF-Hydro outperform the HD model.***

References:

Cisterna-García, A., González-Vidal, A., Martínez-Ibarra, A., Ye, Y., Guillén-Teruel, A., Bernal-Escobedo, L., and Skarmeta, A. F.: Artificial intelligence for streamflow prediction in river basins: a use case in Mar Menor, Sci Rep, 15, 19481, https://doi.org/10.1038/s41598-025-04524-0, 2025.

Hagemann, S. and Dümenil, L.: A parametrization of the lateral waterflow for the global scale, Clim Dyn, 14, 17–31, https://doi.org/10.1007/s003820050205, 1997.

Stacke, T. and Hagemann, S.: HydroPy (v1.0): a new global hydrology model written in Python, Geosci Model Dev, 14, 7795–7816, https://doi.org/10.5194/gmd-14-7795-2021, 2021.

Ruti, P. M., Somot, S., Giorgi, F., Dubois, C., Flaounas, E., Obermann, A., Dell'Aquila, A., Pisacane, G., Harzallah, A., Lombardi, E., Ahrens, B., Akhtar, N., Alias, A., Arsouze, T., Aznar, R., Bastin, S., Bartholy, J., Béranger, K., Beuvier, J., Bouffies-Cloché, S., Brauch, J., Cabos, W., Calmanti, S., Calvet, J.-C., Carillo, A., Conte, D., Coppola, E., Djurdjevic, V., Drobinski, P., Elizalde-Arellano, A., Gaertner, M., Galàn, P., Gallardo, C., Gualdi, S., Goncalves, M., Jorba, O., Jordà, G., L'Heveder, B., Lebeaupin-Brossier, C., Li, L., Liguori, G., Lionello, P., Maciàs, D., Nabat, P., Önol, B., Raikovic, B., Ramage, K., Sevault, F., Sannino, G., Struglia, M. V, Sanna, A., Torma, C., and Vervatis, V.: Med-CORDEX Initiative for Mediterranean Climate Studies, Bull Am Meteorol Soc, 97, 1187–1208, https://doi.org/10.1175/BAMS-D-14-00176.1, 2016.