Anonymous Referee #

With their paper, Galanaki et al. perform a calibration and validation exercise of the fully coupled WRF-Hydro modelling system over the Attica Region, the most densely populated of Greece, considering 7 high rainfall events from 2011 to 2014. Even though the topic addressed is undoubtedly very interesting (an attempt to perform a complete meteo-hydrological forecast over small catchments in a densely urbanized area), my opinion is that, at least at this stage, the paper does not provide new insights, neither concerning methodology (for which I have some concerns) nor regarding results. The most important novelty, according to authors' words, is that "this outcome is important because WRF-Hydro is implemented under calibration with ground-truth observations for the first time in Greece", but in my opinion, it's not enough (otherwise, any first application of WRF-Hydro around the world should deserve publication). I've some major comments and several minor comments listed below. My general opinion is that the paper should be strengthened significantly before being ready to publication, even though I acknowledge that some results if presented better and with more details, could be useful and add information to the topic of fully coupled atmospheric-hydrological modelling and its operational application over small catchments. I hope my comments can help with strengthening the study.

Main comments

Introduction: a lot of work made on meteo-hydrological forecasting chains in the Mediterranean area (and in Greece), even using the WRF-Hydro modelling system, has been not considered, but it should. Please find at the end of the review only a partial list of possible references to be considered.

More studies related to numerical hydrometeorological research has been cited in the Introduction Sections (lines 67-71).

Lines 67-71:

"…The WRF-Hydro model has been used in numerous flood-related research applications (Senatore et al., 2020; Papaioannou et al., 2019; Varlas et al., 2019; Avolio et al., 2019; Lin et al., 2018; Silver et al., 2017; Xiang et al. 2017; Arnault et al., 2016; Givati et al., 2016; Wagner et al., 2016; Senatore et al., 2015; Yucel et al., 2015) and for operational flood forecasting in the United States (Krajewski et al., 2017; NOAA, 2016) and Israel (Givati and Sapir, 2014)."

Calibration methods: I've several concerns. Mainly, it's not clear what is the input precipitation for the calibration of the hydrological model (I wonder if the whole fully coupled system was calibrated upon observed discharge). Furthermore, I've doubts about the final choice of the parameters, which not seldom are equal to one of the limits of the range of scaling factors. I also have other doubts for which I ask the authors to refer to my specific comments. Furthermore, I allow myself to suggest authors read the recently accepted paper of Fersch et al. (2020) dealing in the detail with WRF-Hydro calibration issues.

Concerning the precipitation:

The calibration of the WRF-Hydro model was performed based on the WRF atmospheric forcing, including the precipitation fields. Several preliminary tests have been performed concerning the WRF model configuration (spin-up, physics parameterization; lines 175-177 and 199-204) in order to achieve the most accurate representation of the observed precipitation which is of great importance for simulating the corresponding observed discharge. Corrections have been applied to the manuscript to clarify the above (lines 261-263).

It is worth mentioning that previous studies calibrated the WRF-Hydro model following the same approach of forcing the model with WRF data (e.g., Li et al. 2020; Liu et al., 2020; Li et al. 2017; Silver et al., 2017).

Lines 261-263:

"…The calibration of the WRF-Hydro model was performed using the WRF atmospheric forcing, including the precipitation fields, following the same approach of forcing the model with WRF data from previous studies (e.g. Li et al. 2020; Liu et al., 2020; Li et al. 2017)."

Concerning the calibrated parameters:

The reviewer is right. The manuscript was modified to highlight this fact (lines 300-301 and 338-339).

Lines 300-301:

"…It should be noted that the optimal parameters for REFKDT and RETDEPRTFAC hit the lower and calibration limit, respectively. Relaxing their constraints may result to better calibrations results."

Lines 338-339:

"…As in the case of Sarantapotamos, the optimum value for REFKDT reaches the lower calibration limit indicating that changing the calibration limit may let to better result."

Results: I wonder about the differences between precipitation results with and without fully coupling. Several studies show that for short simulations such as those performed in this study it is very difficult that differences emerge in the precipitation fields due to the differences in soil moisture conditions. Among them, Avolio et al. (2019), which for a case study rather similar to those analyzed by the authors found that correct SST representation is much more impacting. Therefore, more details should be provided by the authors about how they reached their results, and they should try to explain the reasons they got these results.

The differences in the simulated precipitation between WRF-only and WRF-Hydro models have been addressed by examining the soil moisture and latent heat flux before the initiation of the precipitation for each event. Slight differences between the average values of the aforementioned parameters were found, which may affect the resulted precipitation. The authors are aware that this outcome is an indication, as highlighted in the manuscript, and that the effects of soil moisture on precipitation fields are more evident in long-term simulations, when the land surface variables read a steady state (e.g., Senatore et al., 2015). For this, they intend to perform an in-depth analysis for assessing the model's surface energy budget in a follow-up study.
The manuscript was modified to clarify the above (lines 388-399 and 436-447)

Lines 388-399:

"…Table 7 shows the basin average soil moisture (at the 1st level) and latent heat flux simulated by the WRF-Hydro and WRF-only models, at the time before the beginning of the examined storms events. As can be seen the soil moisture differences between the models range from 0.005 to 0.027 $m^3\ m^{-3}$ and latent heat flux differences span from 0.038 to 16.862 W/m^2. These differences simulated by the two models provides an indication that the most accurate replication of the observed precipitation provided by the WRF-Hydro model compared to the WRF-only model is related to the physical process associated with the coupling of land-atmosphere and hydrological routing in the WRF-Hydro model. In particular, WRF-Hydro, affects the soil moisture content, due to the computation of the lateral redistribution and re-infiltration of the water (Gochis et al., 2013), which in turn influences the computation of the sensible and latent heat fluxes. These fluxes are associated with humidity and temperature in the lower atmosphere and consequently precipitation (Seneviratne et al., 2010). However, it should be noted that the effects of soil moisture on precipitation fields are more evident and valid in long-term simulations when the land surface variables reach a steady state (Fersch et al., 2020; Senatore et al., 2015)."

Lines 436-447:

"A preliminary analysis of key water budget components indicated that the precipitation simulation improvement provided by the WRF-Hydro system may be related to the feedback of the terrestrial hydrology parameterization on the modeled atmosphere. A follow up study could focus on the further investigation of impact of the more detailed representation of the interaction between the land surface and hydrology processes to the surface energy budget under the WRF-Hydro coupling scheme by applying long-term simulations and validated the results against ground-based or satellite observation, considering limitations arising from internal model variability (Bassett et al., 2020) and domain size (Fersch et al, 2020; Arnault et al., 2018). Also, the incorporation of the SST update into the model will be considered as previous studies shown a positive feedback to simulations (Avolio et al., 2019; Senatore et al., 2015).

Even though a more detailed analysis is required to explore the sensitivity of the simulated precipitation to the coupling between hydrological and land-atmosphere processes, the current study demonstrates that the coupled WRF-Hydro model has the potential to enhance precipitation forecast skill for operational flood predictions."

Furthermore, concerning the presentation of the results themselves, much more details should be given (please refer to specific comments).

Please find the author's responses in the specific comments.

Concerning the utility of the study for "operational forecasting purposes", the authors should at least discuss: 1) why they use in their study reanalyses instead of operational GCM forecasts, which makes their study not completely indicative for operational purposes in terms of forecasts performance; 2) what is the additional computational burden of fully coupled simulations and if it's worth it.

1) Unfortunately, the on-line availability of the GFS forecasts is limited for historical periods as the studied one (2011- 2014). GFS initialization data could be ordered for the investigated events but at a coarse spatial resolution (0.5ºx0.5º), which was not consider adequate for forcing the WRF simulations having a coarse domain (do1) resolution of 18 km. For this, the ERA5 reanalysis data were preferred over the GFS operational forecasts in this study. Concerning the ECMWF IFS forecasts, unfortunately, their availability is restricted to National meteorological services or users with a special paid contract. The manuscript has been modified accordingly (lines 195-198)

Lines 195-198:

"…It should be noted that the use of ERA5 reanalysis data was preferred instead of the operational GFS data, as the on-line availability of the GFS forecasts is limited for historical periods. GFS initialization data could be ordered for the investigated events but at a high spatial resolution of 0.5° × 0.5°, which was not considered adequate for forcing the WRF simulations having a coarse domain (do1) resolution of 18 km."

2) The manuscript has been modified accordingly to address the computation burden of fully coupled simulations (lines 448-454)

Lines 448-453:

"…For an operational point of view, the application of a coupled WRF-Hydro model to exploit its beneficial impact in simulating precipitation is partially limited due to the additional computational time needed for the execution of the WRF-Hydro model. In particular, in our case, a three day coupled WRF-Hydro forecast considering a prior 12 hours spin up under the investigated configuration requires x1.35 time compares to WRF-only implementation in 140 computing nodes. It should be noted that the extra computational time depends on the WRF-Hydro configuration and the computing resources, in which the model is applied."

Finally, I suggest a general review of the text concerning English grammar and style (some comments, as examples, are provided below).

Revisions concerning the English were made throughout the whole manuscript.

Specific and minor comments:

L53: Wagner

Changed accordingly.

Fig. 1a: the hydrological features are not clear. I suggest separate panels where the analyzed catchments (including their borders) are represented better. I guess that, given the high urbanization level, land cover is also an important piece of information to highlight. Finally, all the toponyms cited in the text (e.g., Cithaeron mountain range, Halandri's stream, etc.) should be reported in the map

Fig. 1 was updated accordingly.

L78: increased concerning what? To the past? What period? Please specify, otherwise, I suggest another term (e.g., high?). Anyway, the sentence looks a bit redundant.

The sentence was corrected.

L95: by the Ymittos Mountain

Corrected.

L100: I guess "were provided". This term "provide" is used 4 times in 5 consecutive lines. Probably the text could be revised

Lines 117-126 were modified to address this issue.

L106: I would organize Table 1 from the oldest to the most recent event. Furthermore, I suggest dealing with events #5 and #6 merging them, I guess they depend on the same synoptic situation.

Table has been organized according to the reviewer's suggestion. The old events #1 and #7 have been merged (new event #4), while the old events #5 (new event #2) and #6 (new event #3) were kept separately as they refer to different dates, and, consequently, they are characterized by different atmospheric conditions (lines 127-154).

L114: "were occurred" not correct

Changed to "were reported"

L128: D04

Corrected.

L137: please revise the text

The text was revised.

LL139-147: this information should be included in Table 2, possibly along with the corresponding WRF options

Table 2 was updated accordingly.

L145: it would be useful to explain why the Noah LSM scheme is preferred to the more recent Noah-MP

The manuscript was modified to justify the use of the Noah LSM (lines 185-189).

Lines 185-189:

"...Noah-MP introduces multiple options and tunable parameters to simulate the land surface processes. However, the default values of these options and parameters are not suitable for every study area (e.g. Giannaros et al., 2019). In contrast, the Noah LSM has been tested and applied successfully in several studies focusing in Greece (e.g. Varlas et al., 2019; Papaioannou et al., 2019; Giannaros et al., 2020)."

L157: “The simulation periods for each event are presented in Table 1.” Not clear: do the simulations include always the whole days (i.e., from 00:00 to 00:00)? Anyway, what spin-up times were selected?

The spin-up time and the exact time of the simulations’ start and end are now included in Table 1.

Section 2.2.2. Even if it is already specified in the title of Section 2.2, I would specify here that WRF-Hydro is used in fully coupled (i.e., two way) mode.

The manuscript was changed accordingly.

L167: 605/95 = circa 7. So, the disaggregation factor is 7? Please highlight more this feature and explain your choice.

More information was added concerning the choice of disaggregation factor (lines 215-219).

Lines 215-219:

“…The catchments' routing grids were computed based on SRTM 90 m topography data using the WRF-Hydro GIS pre-processing toolkit. In order to exploit this high-resolution input dataset, avoiding interpolation to a coarser grid (Verri et al., 2017; Gochis and Chen, 2003), a ~95 m spatial resolution WRF-Hydro domain was configured over the WRF innermost domain. Thus, the ratio between the high-resolution terrain routing grid and the WRF land surface model (aggregation factor; AGGFACTRT) was set to 7.”

L183: I’m not aware that the stepwise approach is somehow recommended. There are many examples of mixed or automated calibration approaches. Among the others, I suggest a very recent one by Fersch et al. (2020). The cited work of Cuntz et al. refers to Noah-MP, not to WRF-Hydro.

The reviewer is right. The manuscript was modified accordingly.

L196: I guess “when a parameter was calibrated”

Corrected.

L196: I understand that there’s a kind of hierarchy in parameters calibration, but it’s not clear which is the parameter calibrated first and which later

The manuscript was modified to clarify this issue (lines 252-253).

Lines 252-253:

"…Thus, the parameters were calibrated in the following order: REFKDT, RETDEPRTFAC, OVROUGHRTAC and MannN."

Section 3.1.1: the fundamental information about the initial value of all the calibrated parameters is missing. Furthermore, other information is missing: e.g., what precipitation values were used for the calibration?

Table 3 has been updated to include the default values of the calibrated parameters.
Concerning precipitation, please refer to the main comment concerning calibrated methods.

L217: the value is at the border of the calibration range. This means that probably the authors should explore other lower values for REFKDT, relaxing their constraints. The same for RETDEPRTFAC

Please refer to the main comment concerning calibrated methods.

L219: it's even more unclear what precipitation was used for calibration. I hope observed, not simulated (in Fig. 2 there are two simulated precipitation series)
L224: no displacement would have been necessary if observations were considered.

Please refer to the main comment concerning calibrated methods.

Figs.2, 5, 6, etc. show both WRF-Hydro and WRF precipitations, but they are not introduced and the difference is not explained in due time into the text.

The authors consider essential the fields of observed and simulated (WRF-Hydro) temporal evolution of precipitation to be in the same subplot with the observed and simulated temporal evolution of discharge. Indeed, the discussion concerning the simulated temporal evolution of precipitation from WRF-only simulations is introduced at the last section of the results. We could extract the field of precipitation from WRF-only simulations from the existed figures and reproduce the same figures for the precipitation at the sector 3.3, but we consider that it will be confusing to show these figures twice.

L245: Figs. 5a and 6a refer to precipitation

Corrected.

L248: time of maximum occurrence?

Corrected.

L251: "time of maximum values": not much better definition than before

Corrected.

Section 3.2: for Rafina catchment, same problems as for the previous calibration procedure (please refer to my comments above)

The corresponding corrections were applied in the manuscript

Section 3.3: what stations are considered? All? Only Vilia and N. Makri? Not clear. If it's only Vilia and N. Makri, how were the other stations shown in fig. 1 used?

The analysis was performed using only the stations of Vilia and N. Makri. Corrections have been applied in the manuscript to clarify this fact.
The remaining stations in the old Fig.1 have been utilized in the initial sensitivity tests for finding the best configuration of WRF, the result of which are not included in the manuscript. Fig. 1 was updated to avoid any misconceptions.

L321: Anyah et al.'s work does not regard WRF-Hydro

Removed.

Conclusions: it looks like a summary. It should be enriched highlighting the strong points of the study

This part of the manuscript were modified. We added additional information related to the water budget analysis and the computational burden of the hydrological analysis.