Anonymous Referee #1

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 56-60).

Lines 56-60:

"…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 166-171 and 144-146) 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 220-222).

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 220-222:

"…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 259-260 and 297-298).

Lines 259-260:

"…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 297-298:

"…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 346-357 and 489-400)

Lines 346-357:

"…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 389-400:

"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 (0.5^ox0.5^o), 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 162-165)

Lines 162-165:

"…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 401-406)

Lines 401-406:

“…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:

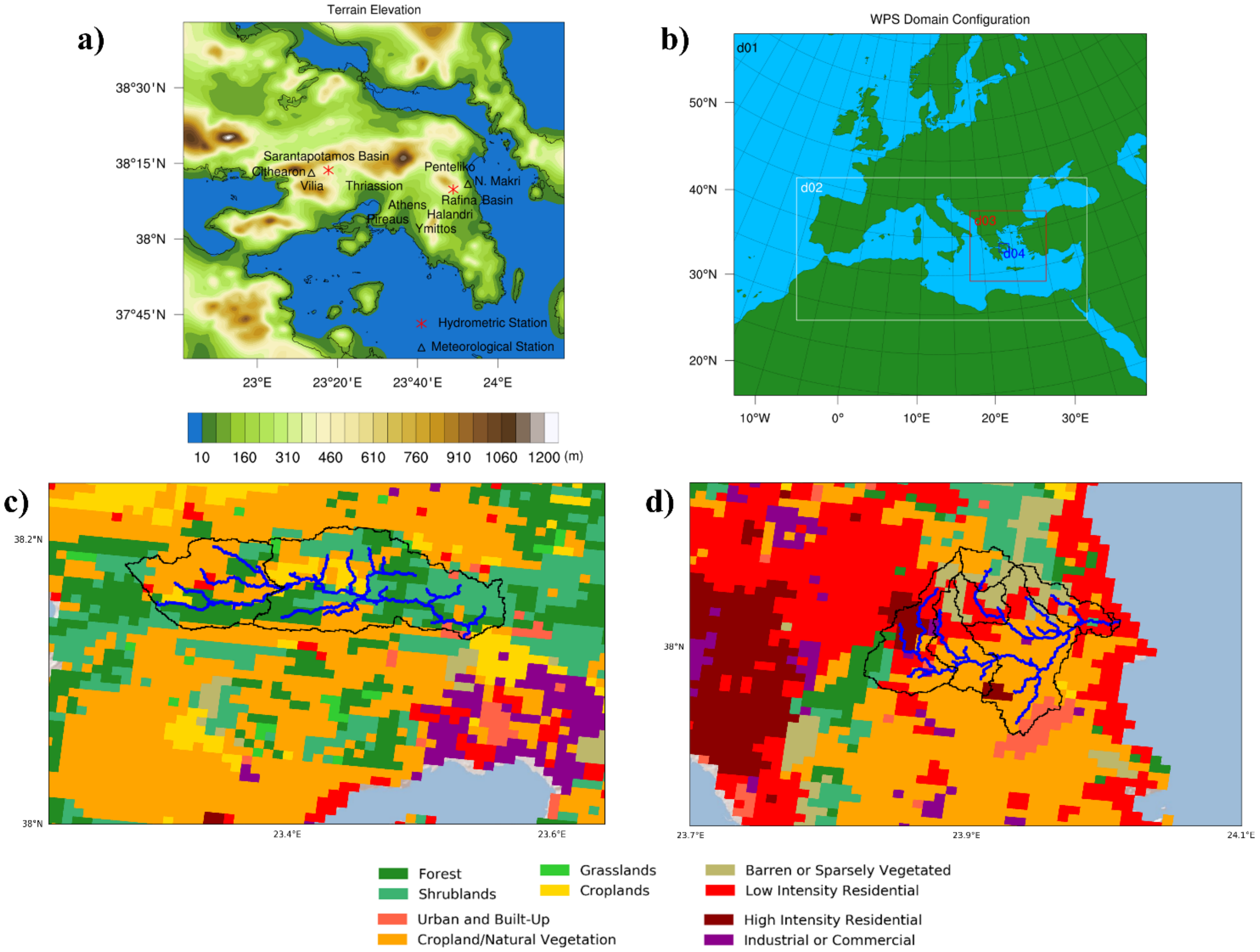
L53: Wagner

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.

Figure 1: (a) Terrain elevation of the studied domain (obtained by MODIS-IGBP global land cover data) along with two channel network and the positions of the meteorological (triangle marker) and hydrometric stations (star marker). (b) Modeling domains. The borders of *analyzed catchments along with the land cover for* (c) Sarantapotamos and (d) Rafina basins



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 (lines 78-79).

Lines 78-79:

"…The population of Attica is ~3.800.000 people (about 36% of the national total) and includes a great part of the national financial and commercial activities…"

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 86-93 were modified to address this issue.

Lines 86-93:

"…In the current study, we focus on two drainage areas of the flood-prone Attica region. The first one is the Sarantapotamos Basin (Figs. 1a, 1c) that drains an area of 310 km^2 and is responsible for flooding events in the urbanized broader area of Thriassion plain, located in west Attica, Greece. Among the most important natural flood causes in the area are the geomorphological characteristics of the drainage network, the intense rainfall and the increasing urbanization which is deprived of integrated flood defense measures. In particular when heavy rainfall occurs, the relatively mild slopes result in a decrease of the surface runoff velocity, accumulating a large volume of water in short times (Zigoura et al., 2014)…"

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 107-125).

Table 1. Simulation periods of each event and hydrometeorological charecteristics

	Basin	Simulation date Start	Simulation date End	Spin-up	Total rainfall	Maximum discharge
Event #1 /E1	Rafina	02/01/2011 00:00 UTC	03/01/2011 18:00 UTC	6h	37.6 mm of rain (24 h accumulated) in N. Makri	8 m^3/s in Rafina
Event #2 /E2	Rafina	02/02/2011 00:00 UTC	05/02/2011 18:00 UTC	24h	123.8 mm of rain (48 h accumulated) in N. Makri	24.3 m^3/s in Rafina
Event #3 /E3	Rafina	06/02/2012 06:00 UTC	08/02/2012 18:00 UTC	6h	33.6 mm of rain (48 h accumulated) in N. Makri	9.1 m^3/s in Rafina

Event #4 /E4R	Rafina	28/12/2012 06:00 UTC	31/12/2012 18:00 UTC	18h	86.8 mm of rain (72 h accumulated) in N. Makri	44.3 m^3/s in Rafina
Event #4/E4S	Sarantapotamos	28/12/2012 18:00 UTC	01/01/2013 18:00 UTC	18h	104.6 mm of rain (72 h accumulated) in Vilia	12.8 m^3/s in Vilia
Event #5 /E5	Sarantapotamos	21/02/2013 18:00 UTC	23/02/2013 18:00 UTC	6h	77 mm of rain (24 h accumulated) in Vilia	19.2 m^3/s in Vilia
Event #6 /E6	Sarantapotamos	02/03/2014 00:00 UTC	04/03/2014 18:00 UTC	24h	85 mm of rain (48 h accumulated) in Vilia	10.7 m^3/s in Vilia

Lines 107-125:

"…Six flood events have been considered for the analysis. Table 1 includes the simulation periods of each event, which were selected after spin-up sensitivity experiments (section 2.2.1), and their observed total rainfall and maximum discharge as they have been recorded at the meteorological and hydrometric stations. All examined episodes were associated with synoptic atmospheric circulation, driven by low-pressure systems, which, in most cases, were combined with 500-hPa troughs and cut-off lows. In particular, surface low-pressure systems, found west of Greece, affected the country in combination with upper-level cut-off lows on 6 February 2012 (event #3) and 29 December 2012 (event #4). In the course of events #1 and #6, the atmospheric circulation was characterized by troughs in the middle troposphere over Greece, associated with surface cyclones located west of North Italy (event #6) and in the Ionian Sea (event #1). The systems induced considerable precipitation in Greece during the above episodes-resulting to noticeable impacts over the examined basins (Giannaros et al., 2020). The higher impacts in Sarantapotamos catchment were reported in Vilia at the night between 21 and 22 February 2013 (event #5), when

24-h precipitation and maximum discharge reached up to 77 mm and 19.2 m^3/s, respectively. During this episode, a very deep surface low crossed the Mediterranean Sea towards Greece. The system was associated with an upper-level trough having a negatively titled axis (Giannaros et al., 2020). Between 02 and 05 February 2011 (event #2), exceptional atmospheric conditions affected Greece (Giannaros et al., 2020). Significant impacts were evident in Rafina catchment, where the total 48-h rainfall surpassed 123 mm in N. Makri and the maximum discharge exceeded 24 m^3/s in Rafina. As highlighted above, the events #2 and 5 affected the examined areas more severely and were the most devastating for the whole area of Attica, where floods, deaths, destruction and great economic losses were induced. More details on the hydrometeorological and socio-economic characteristics of events #2 and #5 can be found in Giannaros et al. (2020)…"

L114: "were occurred" not correct

Changed to "were reported"

L128: D04

Corrected.

L137: please revise the text

The text was revised (lines 137-142).

Lines 137–142:

"…Despite, the high spatial resolution of the MODIS-IGBP dataset, it only includes one category for the urban areas. The latter datasets are considered to be inadequate for ultrahigh-resolution (< 1 km) modeling (Giannaros et al., 2018; Nunalee et al., 2015) , which is necessary for hydrometeorological forecasting (e.g., Verri et al., 2017).Thus, the high resolution Shuttle Radar Topography Mission (SRTM) 90 m × 90 m topography data and the 3-arc-sec resolution Corine Land Cover (CLC) dataset were used for a better land use and topography representation in the innermost d04 domain…"

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

Table 2 was updated accordingly.

Table 2. The WRF Physics schemes used

	Europe (d01)	Mediterranean	Greece (d03)	Attica Basin (d04)

		(d02)		
Microphysics	WSM6	WSM6	WSM6	WSM6
Cumulus physics	KF	KF	-	-
Shortwave/longwave radiation physics	RRTMG/RRTMG	RRTMG/RRTMG	RRTMG/RRTMG	RRTMG/RRTMG
Planetary boundary layer physics	MYJ	MYJ	MYJ	MYJ
Surface layer physics	Eta similarity	Eta similarity	Eta similarity	Eta similarity
Land surface model	Noah	Noah	Noah	Noah

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 152-157).

Lines 152-157:

"...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.

Table 1. Simulation periods of each event and hydrometeorological charecteristics

	Basin	Simulation date Start	Simulation date End	Spin-up	Total rainfall	Maximum discharge
Event #1 /E1	Rafina	02/01/2011 00:00 UTC	03/01/2011 18:00 UTC	6h	37.6 mm of rain (24 h	8 m^3/s in Rafina

					accumulated) in N. Makri	
Event #2 /E2	Rafina	02/02/2011 00:00 UTC	05/02/2011 18:00 UTC	24h	123.8 mm of rain (48 h accumulated) in N. Makri	24.3 m^3/s in Rafina
Event #3 /E3	Rafina	06/02/2012 06:00 UTC	08/02/2012 18:00 UTC	6h	33.6 mm of rain (48 h accumulated) in N. Makri	9.1 m^3/s in Rafina
Event #4 /E4R	Rafina	28/12/2012 06:00 UTC	31/12/2012 18:00 UTC	18h	86.8 mm of rain (72 h accumulated) in N. Makri	44.3 m^3/s in Rafina
Event #4/E4S	Sarantapotamos	28/12/2012 18:00 UTC	01/01/2013 18:00 UTC	18h	104.6 mm of rain (72 h accumulated) in Vilia	12.8 m^3/s in Vilia
Event #5 /E5	Sarantapotamos	21/02/2013 18:00 UTC	23/02/2013 18:00 UTC	6h	77 mm of rain (24 h accumulated) in Vilia	19.2 m^3/s in Vilia
Event #6 /E6	Sarantapotamos	02/03/2014 00:00 UTC	04/03/2014 18:00 UTC	24h	85 mm of rain (48 h accumulated) in Vilia	10.7 m^3/s in Vilia

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 (lines 174-175)

Lines 174-175:

"…The WRF-Hydro modeling system, version 3.0, was used for this study under a fully coupled mode. WRF-Hydro is a distributed hydrometeorological modeling system which is two-way coupled with WRF…"

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 181-185).

Lines 181-185:

"…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 (lines 197-201).

Lines 197-201:

"…Generally, the calibration processes for WRF-Hydro can be divided into three categories: the manual step-wise (e.g. Li et al., 2017), the automate calibration process and mixed calibration approaches combining manual and automate calibration (e.g. Verri et al., 2017). The step-wise approach of calibration is widely applied in order to minimize the high number of model runs and are required for the automate calibration approach…"

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 211-212).

Lines 211-212:

"…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.

Table 3. The range of calibrated parameters

Parameter	Definition	Range of scaling factor	Increment	Default value
REFKDT	runoff infiltration	0.5-1.5	0.1	3.0
RETDEPRTFAC	surface retention depth	0-10	1	1.0
OVROUGHRTAC	surface roughness	0.1-1	0.1	1.0
Manning's roughness/ stream order 1	channel roughness	0.33-1.16	0.1	0.55
Manning's roughness/ stream order 2	channel roughness	0.21-0.74	0.1	0.35
Manning's roughness/ stream order 3	channel roughness	0.09-0.32	0.1	0.15
Manning's roughness/ stream order 4	channel roughness	0.06-0.21	0.1	0.10

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 (lines 286-301).
Lines 286-301
"…The stepwise calibration method suggested above, was implemented for the calibration of Rafina basin using event #2. Figure 6 shows the temporal distribution of the precipitation as observed at the station of N. Makri and simulated using WRF atmospheric only simulations and WRF-Hydro coupled simulations, while Fig.7 shows the temporal evolution of the observed and simulated discharges for the possible values of each calibrated parameter. The observed and

simulated precipitation (provided by WRF-Hydro) are highly correlated (correlation coefficient: 0.83) while quantitatively they also compare very well (Fig. 6). The choice of the optimum values for each parameter was based on the visual comparison of the simulated and observed discharge (Fig. 7) and statistical analysis (Table 5), as it was explained for Sarantapotamos basin.

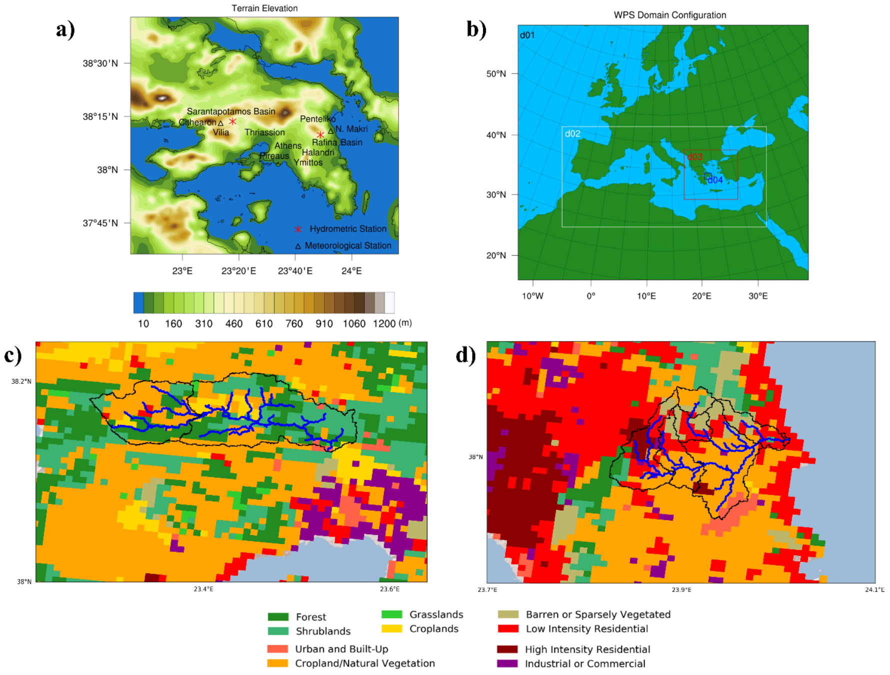
In consistency to the calibration of Sarantapotamos, we firstly performed several simulations for possible REFKDT's values between 1 and 5 and we also found that the appropriate range of the scaling factor from 0.5 to 1.5. Thus, the additional simulations were performed within this range with increment of 0.1 and the value of 0.5 was selected as optimum value for REFKDT parameter. 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. The simulations for RETDEPRTFAC were performed within the range from 0 to 10, with increment of 1. As in the case of Sarantapotamos, the simulated discharge is decreasing with increasing values of RETDEPRTFAC (Fig. 7b). After the comparison of the aforementioned statistical criteria, the selected optimum value for the RETDEPRTFAC parameter was 6…"

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.

Figure 1: (a) Terrain elevation of the studied domain (obtained by MODIS-IGBP global land cover data) along with two channel network and the positions of the meteorological (triangle marker) and hydrometric stations (star marker). (b) Modeling domains. The borders of *analyzed catchments along with the land cover for* (c) Sarantapotamos and (d) Rafina basins



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 was modified. We added additional information related to the water budget analysis and the computational burden of the hydrological analysis (lines 389- 406**)**.

Lines 389- 406

"…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.

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…"

Anonymous Referee #2

General comment

The coupling of land and atmospheric processes and evaluating the impact on the forecast skill compared to atmosphere-only modeling is an important topic for the community and particularly NHESS readers. The manuscript aims to (1) to investigate the ability of WRF-Hydro to simulate selected cases of flood occurrence in the area of Attica (Greece) and (2) to study the influence of land-atmosphere interactions on the improvement of precipitation forecasting. While the first objective is an important effort towards local operational flood forecasting, the second objective would be the main source of novelty and new insights for the scientific community. However, the current version of the manuscript does not thoroughly address this objective and fails to diagnose the physical mechanism explaining the reported improvement from the coupling. My suggestion would be a re-submission after the authors make the below major improvement which may/may not alter the main conclusions of the study.

Major comments

Comment 1:

In order to take the full advantage of the WRF-Hydro system, diagnoses of the feedback processes /mechanisms controlling the water cycle (e.g. runoff, penetration, evaporative fraction, water vapor flux) should be conducted. Such diagnoses may lead to valuable generic outcome that could benefit the research community. The primary mechanism to diagnose is the soil moisture-precipitation feedback loop (El Tahir et al., 1998) and the evolution of surface fluxes during the simulations (uncoupled vs. coupled) – see for example the recent works of Kumar et al. (2020) and Wehbe et al (2019). It is strongly recommended that such diagnoses are explored to confirm speculative statements, such as that mentioned in Line 302: “The improved simulation of the soil moisture affects the computation of the sensible and latent heat fluxes, which influence humidity and temperature in the lower atmosphere and consequently precipitation. Therefore, the physical process of the coupling of land-atmosphere is expected to improve the forecast skill of precipitation”.

Comment 2:

Please specify if a two-way or one-way grid nesting was employed. This is a crucial point.

If a one-way grid nesting was used, the authors have to make sure that domains 1, 2 and 3 are identical in both WRF and WRF-Hydro simulations. This may not be the case if the authors used two different executables, one for WRF and the other for WRF-Hydro. If domains 1, 2 and 3 in the

WRF and WRF-Hydro simulations are different, then it can be argued that the differences obtained in domain 4 are not due to the consideration of lateral hydrological processes, but to different large-scale forcing. In this case the main conclusion of the paper has to be revised.

If a two-way grid nesting was used, then the above effect is masked by the feedbacks from domain 4, which are unlikely to be exactly the same between the WRF and WRF-Hydro simulations. Still, the fact that domain 1, 2 and 3 would be different in this case would not be necessarily due to the feedbacks from the resolved lateral water flow in domain 4, but simply internal atmospheric variability. The authors are very quick in concluding that the improved precipitation in the WRF-Hydro simulation is due to the coupling with lateral terrestrial hydrological processes, which is then taken for granted through the rest of the manuscript. But in my opinion, this improvement would rather be due to atmospheric internal variability, which is a well-known limitation of regional atmospheric models (e.g. Rassmussen et al. 2012).

So in any case the authors have to provide an estimation of internal atmospheric variability, in order to prove that the claimed improvement in modeled precipitation with WRF-Hydro is not the result of a random realization of the considered atmospheric situation. In other words, the authors have to provide an ensemble and assess the robustness of a potential improvement with WRF-Hydro. The ensemble could be generated, for example, by disturbing the initial condition, or by using the GEFS ensemble forecast runs. This ensemble could simply be generated, for example, by adding random perturbation in the soil moisture initial condition, or whatever prognostic variable.

Concerning the reviewer‘s suggestions in main comments 1 and 2:

Indeed, taking full advantage of a two-way coupled hydrometeorological model requires assessing its ability to improve the physical realism concerning land-atmosphere and hydrological interactions, and their impact on precipitation. Such an assessment is more relevant to long-term simulations, when the land surface variables reach a steady state and affect more evidently the precipitation formation (e.g., Senatore et al., 2015). Also, the authors acknowledge that internal model variability (IVM) is an important issue concerning regional atmospheric models (e.g., Bassett et al., 2020). However, both the detailed analysis of the model’s water and energy budget and the investigation of uncertainties arising from IVM are out of the scope of the preset study.

The current study aims principally on assessing the capability of the coupled WRF-Hydro model as an operational short-term flood forecasting system, as given the susceptibility of the study area (Attica) to flooding, which is sufficiently described in the introduction, the development of such an operational tool is considered of great importance. In this framework, the study also investigated the impact of applying a coupled hydrometeorological model on the precipitation forecast skill. The results showed that the coupled WRF-Hydro model has the potential to improve the precipitation forecast accuracy, which is essential for flood forecasting purposes. Following the reviewer's suggestion, a **preliminary analysis was added to the manuscript regarding key water budget components**, indicating 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. The authors acknowledge that this outcome is just an indication and that a more detailed analysis is required to confirm this. Recognizing the importance of such an in-depth analysis, the authors intend to perform it in the future as a follow-up study, considering limitations arising from IVM.

The manuscript was modified to clarify the above (lines 346-357, 389-400), as well as the nesting approach applied for the simulations (lines 129-132)

Lines 346-357:

"…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 0.0269 $m^3\ m^{-3}$ and latent differences span from 0.0376 to 16.8621 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 389-400:

"A preliminary analysis of key water budget components indicated that the precipitation simulation improvement provided by the WRF-Hydro system may 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 consider 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."

Lines 129-132:

"…The Advanced Research Weather Research and Forecasting model Version 3.9.1.1 was used in this study (Skamarock et al., 2008) for the land-atmosphere simulations which were carried out using four two-way nested grids (Fig. 1b): d01, d02, d03 d04 with 18 km (325 × 285 grid points), 6 km (685 × 337 grid points), 2 km (538 × 499 grid points) and 667 m (208 × 184 grid points) grid increments, respectively."

Comment 3:

Why was event #2 selected for the calibration among the other events? Please add more details on the structure/scale of these events – were they all microscale, mesoscale or synoptic situations? This has severe implications on the robustness of the conclusions which may be governed by the microphysics options rather than the WRF-Hydro coupling. The authors select the WSM6 microphysics scheme without providing any justification. Are their previous sensitivity studies done for Greece or the surrounding region to support this selection and its relevance to the simulated storm scale(s)?

Concerning the events:

The selection of events #2 and #5 is primary related to the capability of the model to reproduce the observed rainfall in the study catchments, as an accurate representation of the atmospheric forcing is important for the simulation of the stream discharges and, consequently, for the calibration process.

The description of the synoptic conditions related to the examined events has been updated in lines 107-125.

Lines 107-125:

"…Six flood events have been considered for the analysis. Table 1 includes the simulation periods of each event, which were selected after spin-up sensitivity experiments (section 2.2.1), and their observed total rainfall and maximum discharge as they have been recorded at the meteorological and hydrometric stations. All examined episodes were associated with synoptic atmospheric circulation, driven by low-pressure systems, which, in most cases, were combined with 500-hPa troughs and cut-off lows. In particular, surface low-pressure systems, found west of Greece, affected the country in combination with upper-level cut-off lows on 6 February 2012 (event #3) and 29 December 2012 (event #4). In the course of events 1 and #6, the atmospheric circulation was characterized by troughs in the middle troposphere over Greece, associated with surface cyclones located west of North Italy (event #6) and in the Ionian Sea (event #1). The systems induced considerable precipitation in Greece during the above episodes resulting to noticeable impacts over the examined basins (Giannaros et al., 2020). The higher impacts in Sarantapotamos catchment were reported in Vilia at the night between 21 and 22 February 2013 (event #5), when 24-h precipitation and maximum discharge reached up to 77 mm and 19.2 m^3/s, respectively. During this episode, a very deep surface low crossed the Mediterranean Sea towards Greece. The system was associated with an upper-level trough having a negatively titled axis (Giannaros et al., 2020). Between 02 and 05 February 2011 (event #2), exceptional atmospheric conditions affected Greece (Giannaros et al., 2020). Significant impacts were evident in Rafina catchment where the total 48-h rainfall surpassed 123 mm in N. Makri and the maximum discharge exceeded 24 m^3/s in Rafina. As highlighted above, the events #2 and #5 affected the examined areas more severely and were the most devastating for the whole area of Attica, where floods, deaths, destruction and great economic losses were induced. More details on the hydrometeorological and socio-economic characteristics of events #2 and #5 can be found in Giannaros et al. (2020)."

Concerning the model configuration, several preliminary tests have been performed in the framework of setting up the model for operational forecasting in Greece. The manuscript was modified to clarify the above and justify the selection of physics parameterizations (lines 143-148).

Lines 143-148:

"…The selection of the physics schemes was based on sensitivity tests conducted for the exploration of the best-performing schemes in terms of precipitation forecasting in the framework of setting up the model for operational forecasting in Greece. For the cloud microphysics processes, the WRF Single-Moment 6-Class Microphysics scheme (WSM6; Hong and Lim, 2006) was used, which has been also implemented in other studies over Greece (e.g. Emmanouil et al., 2021; Politi et al., 2018; Giannaros et al., 2016; Pytharoulis et al., 2016)."

Minor comments/corrections

Line 145: please justify the selection of the NOAH LSM instead of the NOAH-MP LSM (also comment on the selection of the MYJ PBL scheme vs. other schemes).

The above suggestions have been applied to the manuscript at lines 152-157.

Lines 152-157:

"…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). In addition, MYJ parameterization scheme has been successfully implemented in other studies over Greece (e.g. Emmanouil et al., 2021; Politi et al., 2018)."

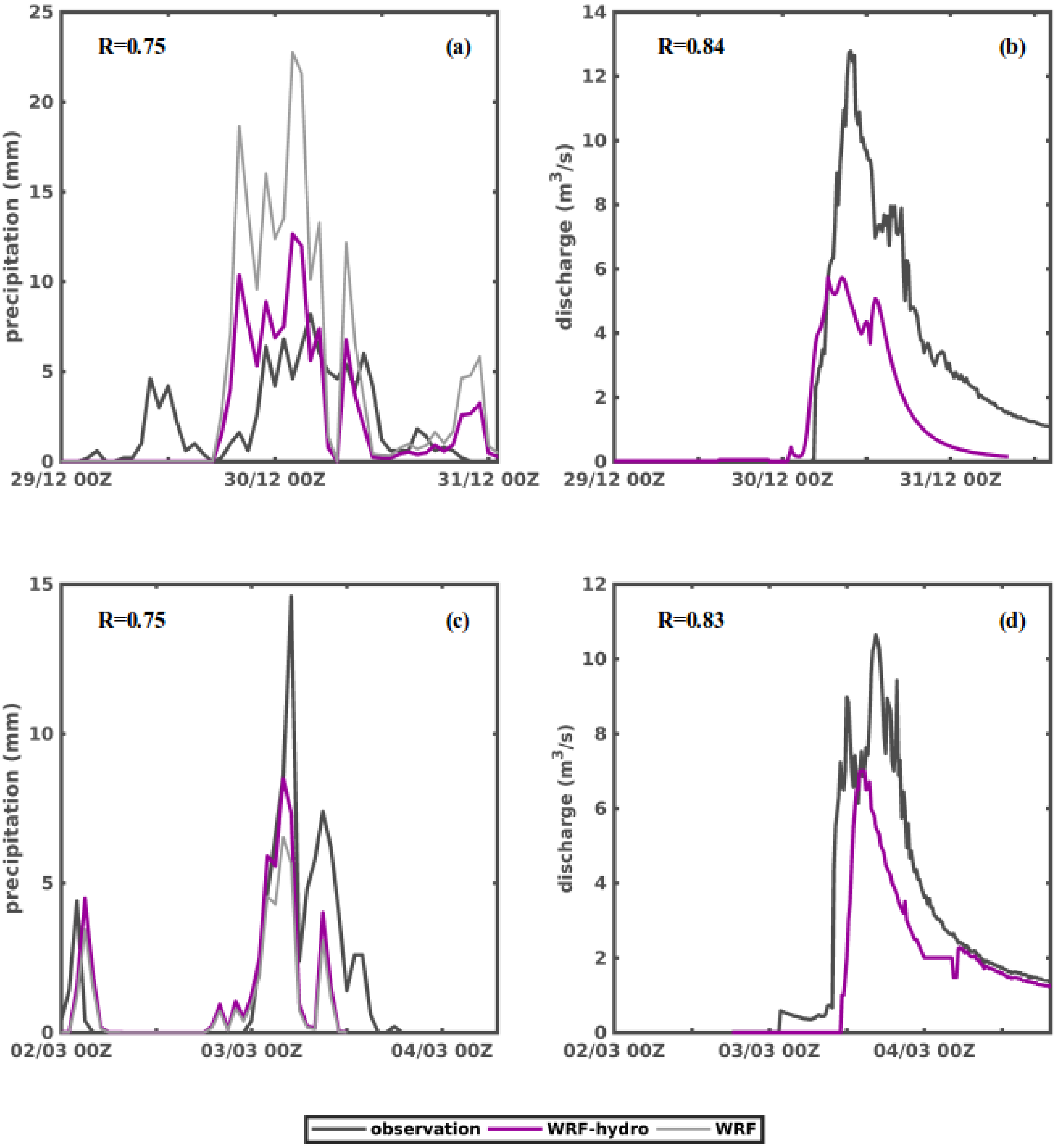
Line 8 (abstract): This study presents an integrated modeling approach for simulating flood events.
Line 12: Remove "on the improvement of"
Line 14: carried out with "the" WRF-Hydro model. There should also be mention of the comparison with WRF-only (standalone/uncoupled) runs.
Line 26: …especially "in its capital, Athens," flooding events…
Line 51: revise to "WRF-Hydro is a recently developed coupled hydrometeorological system that has been used for numerous research applications
Line 61: remove "the" before 36%
Line 75: add "the" before Cithaeron
Line 86: revise to "In the current study, we focus on two…"
Line 89: replace "intense" with "increasing" before urbanization
Line 100-103: capitalize "H" in "WRF-hydro" and correct the sentence structure.
Line 106: "Namely" is used incorrectly here
Line 113: add of: "...the whole of Greece…"
Line 137: add for "…of the area for better simulation…"
Line 218: Use either the long dash (–) or short dash (-) concisely for the term Nash-Sutcliffe

All the above issues have been addressed in the manuscript, as the reviewer suggested

Line 140: please justify the selection of WSM6 MP scheme for the study domain. Are their sensitivity studies done for Greece or the surrounding region to support this selection?
Please refer to the main comment concerning the microphysics scheme.

Figures:
Merge figures 5 and 6 using subplots and add error metrics on each subplot
Merge figures 9, 10 and 11 using subplots and add error metrics on each subplot
We have modified the figures 5,6, 9,10 and 11 according to the suggestions

The figures 5 and 6:



The figures 9, 10 and 11: