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 387-399, 436-447), as well as the nesting approach applied for the simulations (lines 158-162)

Lines 387-399:

"…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 0.0269 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 436-447:

"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 158-162:
"…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 127-154.
Lines 127-154:
"…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 175-181)

Lines 175-181:

"…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 185-190.

Lines 185-190:

"…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