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.

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)?

Minor comments/corrections

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 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?

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).

Line 218: Use either the long dash (–) or short dash (-) concisely for the term Nash-Sutcliffe

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

References:

- Eltahir, E. A. (1998). A soil moisture–rainfall feedback mechanism: 1. Theory and observations. *Water resources research*, *34*(4), 765-776.
- Kumar, S., Newman, M., Lawrence, D. M., Lo, M. H., Akula, S., Lan, C. W., ... & Lombardozzi, D. (2020). The GLACE-Hydrology Experiment: Effects of Land–Atmosphere Coupling on Soil Moisture Variability and Predictability. *Journal of Climate*, *33*(15), 6511-6529.
- Wehbe, Y., Temimi, M., Weston, M., Chaouch, N., Branch, O., Schwitalla, T., ... & Al Mandous, A. (2019). Analysis of an extreme weather event in a hyper-arid region using WRF-Hydro coupling, station, and satellite data. *Natural Hazards & Earth System Sciences*, *19*(6).
- Rasmussen, S. H., Christensen, J. H., Drews, M., Gochis, D. J., & Refsgaard, J. C. (2012). Spatial-scale characteristics of precipitation simulated by regional climate models and the implications for hydrological modeling. *Journal of Hydrometeorology*, *13*(6), 1817-1835.