**Anonymous Referee #1**

I thank the authors for carefully considering my comments. I acknowledge that the quality of the paper has improved to a certain extent. However, I still have some doubts, which I express starting from my previous comments.

Introduction: in my opinion, it's not only a matter of listing a series of studies using WRF-Hydro for reproducing/analysing flood events, rather the authors should better contextualize their research in the framework of those applications, particularly in the Mediterranean area. I ask for further effort in providing reasoned literature background.

The introduction was modified in order to highlight the findings of all recent hydro-meteorological studies that used WRF-Hydro model focusing in the Mediterranean area (lines 61-70).

**Lines 61-70:**

"… Especially, in the Mediterranean area, Senatore et al. (2020) implemented WRF-Hydro in a catchment of Italy in order to highlight the impact of SST in operational forecasts. Avolio et al. (2019) showed that WRF-Hydro was capable to simulate the hydro-meteorological impact of a highly intensity rainfall event in Italy. Senatore et al. (2015) studied the impact of fully-coupled WRF-Hydro model in the forecasting precipitation and showed that the coup[led model provides improved simulation precipitation compared to those provided by WRF-only simulations. Furnari et al. (2020) showed that the implementation of WRF-Hydro has the potential to improve by up to 200% the precipitation forecasts over a small catchments area in Italy. Camera et al. (2020) has implemented WRF-Hydro in small catchment areas in Cyprus and showed how the accuracy of the input precipitation can strongly affect the hydrological simulation. Furthermore, Varlas et al. (2019) and Papaioannou et al. (2019) have showed that WRF-Hydro demonstrate adequate skill in reproducing two past flood events in Greece."

Calibration methods: LL 259-60 and 297-98: I still think that the authors should explore what happens if they relax the calibration boundaries.

The selected boundaries for the calibration in this study are the boundaries that have been used in several previous studies available in the literature (e.g. Kerandi et al., 2018; Naabil et al., 2017; Givati et al., 2016; Yucel et al., 2015). The authors acknowledged that a possible relaxation of the boundaries might result to slightly better calibration results but as the calibration process in this study is performed manually, further exploration of the calibration boundaries would require almost 40 additional simulations for each event provided in this paper. Unfortunately the current computational availability does not allow us to perform these experiments. But certainly this is an issue that deserves further study in the future. Your important comment was addressed as follows in the text (lines 311-314):

**Lines 311-314:**

"For the needs of this study the range of 0.5 to 1.5 was used as this same range was proposed in the literature by previous studies (Kerandi et al., 2018; Naabil et al., 2017; Givati et al., 2016;

Yucel et al., 2015). However lower values than the lower boundary of 0.5 might provide improved results and this issue deserves further investigation in the future"

Results:
Though the soil moisture state immediately before the event is important, I think that concerning latent heat flux the analysis should be made in terms of accumulated values (at least, let's say, starting one day before). However, most importantly, the selected events are characterized by a strong shoreward flux of humid air, which most probably almost hide the effect of the land surface as a moisture source. What I mean is that moisture contribution from the land surface during the events is a second-order process, compared to sea surface contribution. I think that the authors should at least discuss this point.

The analysis concerning the heat flux was made in terms of accumulated values, as the reviewer suggested. The temporal period of accumulation was specified 6 hours before the beginning of the event, as for some events the spin-up period was less than 1 day. The manuscript at lines 362-365 and the results at the table 7 were modified accordingly.

**Lines 362-365:**
"…Table 7 shows the basin average soil moisture (at the 1st level) and 6 hour accumulated 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 accumulated latent heat flux differences span from 4.1 to 41.8 $W/m^2$…."

Table 7

| | Basin | | Soil moisture ($m^3 m^{-3}$) | Accumulated latent heat ($W/m^2$) |
|---|---|---|---|---|
| Event #1 /E1 | Rafina | WRF-Hydro | 0.2915 | 1.4 |
| | | WRF | 0.3034 | -2.7 |
| Event #2 /E2 | Rafina | WRF-Hydro | 0.2760 | 40.1 |
| | | WRF | 0.2660 | 39.3 |
| Event #3 /E3 | Rafina | WRF-Hydro | 0.3427 | 388.1 |

| | | | | |
|---|---|---|---|---|
| | | WRF | 0.3159 | 346.3 |
| Event #4 /E4R | Rafina | WRF-Hydro | 0.2126 | -29.3 |
| | | WRF | 0.2121 | -29.1 |
| Event #4/E4S | Sarantapotamos | WRF-Hydro | 0.2248 | 235.2 |
| | | WRF | 0.2316 | 225.7 |
| Event #5 /E5 | Sarantapotamos | WRF-Hydro | 0.2834 | -9.4 |
| | | WRF | 0.2823 | -10.7 |
| Event #6 /E6 | Sarantapotamos | WRF-Hydro | 0.2792 | 20.3 |
| | | WRF | 0.2666 | 10.5 |

Concerning the land-sea interaction the manuscript was modified in order to highlight further this (lines 373-375).

**Lines 373-375:**
"…Furthermore, the soil moisture is strongly dependent to the sea-atmosphere interactions (Avolio et al., 2019; Senatore et al., 2015) and the synoptic scale circulation…"

Concerning the use of ERA5 reanalyses, in principle, they do not resemble either the reliability of 2011-2014 operational forecasts or current operational forecasts. Of course, I don't ask for considering more recent events with higher resolution GFS forecast, however, in my opinion, the authors have two options: 1) using 0.5° resolution GFS forecasts analyzing the "operational forecasting purposes" at that time; 2) using ERA5 reanalyses and re-modulating the discussion acknowledging that the results achieved are only partially informative concerning operational forecasts. A discussion about the difference in accuracy/reliability of reanalyses/real-time operational boundary conditions could provide a more complete picture of the problem.

The presented analysis was performed using ERA5 data as:

1) The available GFS data for this historical period were not adequate for forcing the WRF simulations in terms of spatial resolution
2) We have not access to operational ECMWF IFS forecasts
3) The ERA5 data have been utilized in several hydrological studies, indicating that they are adequate for hydrological modeling applications
Thus, we used the best available data sources in order to make an attempt to build a flood forecasting system based on the WRF-Hydro model.
Currently, the WRF-Hydro forecasting model is implemented in a pre-operational mode forced by GFS operational data (which is now 0.25°). A follow-up study could focus on evaluating the performance of the model during this pre-operational application.
The manuscript has been modified accordingly in order to clearly specify the above points in the methodology (lines 177-178) and in the results (lines 429-434).

**Lines 177-178:**
"…Furthermore, the ERA5 reanalysis dataset has been proved reliable for hydrological modeling applications (Alves et al., 2020; Crossett et al., 2020; Martens et al, 2020; Tarek et al., 2019)…"

**Lines 429-434:**
"…A follow-up study could focus on evaluating the performance of the model initialized by GFS data during a pre-operational application covering a whole hydrological year. Such a study could enhance our knowledge about the added value of the WRF-Hydro coupled mode and shed light on the performance of the model using GFS operational data. The utmost goal is to provide citizens and stakeholders with reliable information and warnings in order to enhance flood risk awareness and protect lives and properties.…"

Conclusions: I suggest highlighting better why the fully-coupled WRF-Hydro operational forecasts should be preferred to the one-way coupled, notwithstanding the increased computational burden. Is the improvement worth it? The authors should make clear their reasoning (e.g., what kind of trade-off between accuracy and the computational burden they consider to get their conclusions).

It should be noted that the current study does not argue that the fully coupled WRF-Hydro should be preferred compared to the one-way coupled WRF-Hydro for operational application.
Our results indicate a beneficial effect of the coupled WRF-Hydro in simulating accurately the precipitation compared to the WRF-only. This is important for the rainfall-dominated catchments as the studied ones, when mostly precipitation drives the hydrological response.

**Anonymous Referee #2**

The authors have addressed all my comments and justified their work as a preliminary attempt to develop a flood forecasting system based on the WRF-Hydro model. I only have a few minor comments for them to address before the manuscript can be accepted.

Line 46: also cite the original work which diagnoses the soil moisture-precipitation feedback by: Eltahir, E. A. (1998). A soil moisture–rainfall feedback mechanism: 1. Theory and observations. Water resources research, 34(4), 765-776.

Line 49: also cite more recent work on simulating extreme events using WRF-Hydro by:
-Wehbe, Y., Temimi, M., Weston, M., Chaouch, N., Branch, O., Schwitalla, T., Wulfmeyer, V., Zhan, X., Liu, J. and Mandous, A.A. (2019). Analysis of an extreme weather event in a hyper-arid region using WRF-Hydro coupling, station, and satellite data. Natural Hazards and Earth System Sciences, 19(6), 1129-1149.

- Pal, S., Dominguez, F., Dillon, M. E., Alvarez, J., Garcia, C. M., Nesbitt, S. W., & Gochis, D. (2020). Hydrometeorological Observations and Modeling of an Extreme Rainfall Event using WRF and WRF-Hydro during the RELAMPAGO Field Campaign in Argentina. Journal of Hydrometeorology.

Line 53: After mentioning WRF model for the first time, cite:
Skamarock, W. C., Klemp, J. B., Dudhia, J., Gill, D. O., Barker, D. M., Wang, W., & Powers, J. G. (2005). A description of the advanced research WRF version 2. National Center For Atmospheric Research Boulder Co Mesoscale and Microscale Meteorology Div.

We would like to thank the reviewer for the suggestions. All the above citations have been properly implemented in the manuscript (lines 44-50 and 52-54).

**Lines 44-50:**

"…The terrestrial hydrological processes affect soil moisture, a variable that is crucial for the computation of the sensible and latent heat fluxes, which in turn affect the atmospheric response (Seneviratne et al., 2010; Maxwell et al., 2007, Etalhir, 1998). Several studies have shown that an improvement, although not always significant, on the forecasting of the spatiotemporal distribution of extreme synoptic and convective precipitation is provided through the use of coupled hydrometeorological models (e.g., Pal et al., 2020; Wehbe et al., 2019; Senatore et al., 2015; Shrestha et al., 2014; Maxwell et al., 2007). .."

**Lines 52-54:**

"…WRF-Hydro, an enhanced version of the Weather Research and Forecasting (WRF; Skamarock et al., 2005) model, is one of the various modeling systems that provides a two-way coupling between the hydrological and land-atmosphere processes…"

Figure 1 caption: part of the text is italic.
The caption was corrected

Figure 9: please improve the quality of the figure and consider using more contrasting colors for WRF vs. WRF-Hydro.

Figure 9 was updated accordingly