Comment on hess-2021-632
Anonymous Referee #1

I have read the paper by Alfieri et al with great interest and I think the authors present a very well-motivated study with hydrological simulations that sustain their objectives. I am impressed by the high-quality EO data that went into the analysis. I have a few comments which I hope can help the authors to make the paper more structured and the results of the hydrological simulations easier to follow.

General comments

I have troubles with the GLEAM data. Please state very explicit if you are using PET (potential ET) or AET (actual ET) and for what purpose. I assume Fig3 shows AET, but as model forcing you likely use PET. The spatial pattern in Fig3 is hard to interpret as I am not sure about what variable is shown. GLEAM estimates AET and I wonder why it has not been included in the data assimilation (or calibration), similar to SM and snow depth? Please be more explicit and state that the GLEAM scenario uses PET from GLEAM. If I understood this correctly.

I would also like to hear the authors reasoning on why snow depth data and SM data was used in a data assimilation framework and not utilized in the calibration instead?

Figure 10: After reading the paper, I am still unsure what forcing data was used for the baseline run. Especially for the PET data I am quite unsure. I have a hard time understanding the last three experiments. For example PE is never defined and I am left guessing what the abbreviation means. If I understand the three last experiments correct, the model has been recalibrated, but why does the calibration performance not improve?

Specific comments

An additional fifth point to the benefits of EO data to enhance hydrological models: Improved parametrizations (Many examples out there... such as: https://doi.org/10.5194/hess-22-1299-2018).

Page 3 line 85. Please explain what is meant by soil capacity.

The Continuum model is mentioned in section 2.1. and 2.2 prior to its introduction.
Page 4 lines 104-109: It seems a lot of work went into processing the meteorological data. Has this been done by the authors? If not please cite relevant literature.

Page 5 line 130. MCM has not been introduced here.

For the verification of E and SM, on pages 7 and 8, R2 has been used as metric. I suggest to also add a RMSE for the verification against station data. For the snow depth verification, no metric is stated. I suggest to add the same metrics here.

Please show an average SM map next to Fig4, just like you already do for the E data.

The spatial resolution of the hydrological model needs to be stated.

Please specify what PET dataset was used in the baseline model.

Please elaborate more on the hydrological model parametrization and processes. For example, how is AET calculated. I also wonder how some of the soil inputs from SoilGrids are turned into model parameters, i.e. where any regionalization functions applied?

Why did the authors choose to calibrate against the KGE of the flow duration curves and not the actual timeseries? Also, please elaborate more on the selectin of the validation sites.

Fig 9 and Fig 11. In my opinion the color bar (white to blue) is not optimal for the KGE. The river network is blue and the map background is white. Please choose something with more contract to the background.

Page 20, lines 388-392: It seems that performance is reported for all stations. In my opinion only the performance for the validation sites should be used reported general model performance.