[Reviewer comments in normal font; *Author replies in itialic*]

REVIEWER 1

The manuscript provides a welcomed intercomparison of approaches and frameworks to quantify irrigation from remotely sensed soil moisture and ET data. Overall, the results improve understanding of the shortcomings and pitfalls of various approaches designed for this task. I recommend publishing the manuscript after addressing the below comments. This study is likely to inform future irrigation quantification analyses.

Reply: *We thank the reviewer for the positive and constructive feedback that will help us to further improve our work. Below, we outline how we consider responding to the issues pointed out by the reviewer in the revision and what changes we intend to implement.*

Primary concern

Eq. 3: This study calculates irrigation as the sum of SM and ET residuals (satellite – model). This seems like irrigation will often be double counted, because surface soil biases propagate to ET biases. There is a time lag, such that a wetter soil surface (from the satellite) that is attributed to irrigation from a previous time step will be accounted for as irrigation again in a later time step when the water returns to the atmosphere. Is it possible that this could be compensating for other sources of error that favor underestimated irrigation?

If I am missing something about this method that is implemented to avoid double counting, please elaborate on this; if not, please discuss this source of error in the manuscript and how it may be data-specific.

Reply: *Thank you for your comment. The SMOS DISPATCH dataset represents the topsoil (upper 5 cm) soil moisture. This information is difficult to use in the calibration of a hydrological model because the obtained soil moisture could have been obtained doing or just after a rainfall event. In this case, most soil water will be in the topsoil, and during the following days, the water will infiltrate deeper into the soil or evapotranspire. On the other hand, if the hydrological model receives the same rainfall input, the model will distribute the added water between all soil layers and outgoing fluxes within a timestep (one day). This makes it challenging to directly compare model estimates and satellite observations of soil moisture. We have tried to overcome this issue by using an equation from Albergel et al 2008 that uses topsoil soil moisture to estimate the amount of water in the root zone which is more comparable with the hydrological model simulations. During this conversion, we have removed a substantial amount (around 40% of the summed soil moisture increase) of the topsoil soil moisture that enters the soil. The amount of water removed is assumed to represent evapotranspiration deeper infiltration and drainage. The conversion to rootzone soil moisture is the main reason why we do not expect the double counting of irrigation water to be a significant part of our results.*
Plan for revision: *We will make sure that this is better described in the manuscript.*

Specific comments

Lines 109, 115: the semicolon should be a colon.

Plan for revision: *Will be changed in the manuscript.*

Line 120: Please add an additional sentence that explains the reasoning/logic for the assumed losses.

Reply: *Thank you for your comment.*
Plan for revision: *We will incorporate these reasons into the manuscript.*

Section 3.1: Can you please provide justification of why MODIS ET is used instead of other products?

Reply: *We choose the MODIS 16 ET product because it was found to provide valid estimates across European catchments compared to other products (Stisen et al., 2021) and to be influenced by irrigation (Dari et al., 2022; Stisen et al., 2021).*
Plan for revision: *Make it clear that the choice of MODIS ET was based on other studies' comparisons of different ET products.*

Section 3.1: Was a downscaled SMOS product used instead of other high-spatial-res produces (e.g., SMAP-S1) to maintain high temporal resolution? Has there been validation to ensure the SMOS SM product maintains irrigation signals (e.g., similar to Lawston et al., 2017 or Jalilvand et al., 2021)? I believe this is important because the Kumar et al. (2015) analysis seems to suggest the SMOS product fails to detect a significant portion of irrigation signals (which is also related to my concern above regarding Eq. 3).

Lawston, P.M., Santanello, J.A., Kumar, S.V., 2017. Irrigation Signals Detected From SMAP Soil Moisture Retrievals: Irrigation Signals Detected From SMAP. Geophys. Res. Lett. 44, 11,860-11,867. https://doi.org/10.1002/2017GL075733

Jalilvand, E., Abolafia-Rosenzweig, R., Tajrishy, M., Das, N.N., 2021. Evaluation of SMAP-Sentinel1 High-Resolution Soil Moisture Data to Detect Irrigation over Agricultural Domain. IEEE J. Sel. Top. Appl. Earth Observations Remote Sensing 1–1. https://doi.org/10.1109/JSTARS.2021.3119228

Kumar, S.V., Peters-Lidard, C.D., Santanello, J.A., Reichle, R.H., Draper, C.S., Koster, R.D., Nearing, G., Jasinski, M.F., 2015. Evaluating the utility of satellite soil moisture retrievals over irrigated areas and the ability of land data assimilation methods to correct for unmodeled processes. Hydrol. Earth Syst. Sci. 19, 4463–4478. https://doi.org/10.5194/hess-19-4463-2015

Reply: *The SMOS-DISPATCH product was used in this area due to its high spatial and temporal resolution and well-documented ability to detect irrigation in this region (Merlin et al. 2013, Dari et al. 2021). However, the original SMOS SM 40 km product does not reflect the spatial distribution of soil moisture from irrigation, meaning that the spatial soil moisture*

irrigation signal comes from DISPATCH downscaling that uses MODIS NDVI and land surface temperature. We have analyzed the SMOS-DISPATCH product by comparing climatologies of irrigated and rainfed cropland to ensure that the areas have similar soil moisture content during winter (no irrigation) and diverging content during the irrigation season as expected.
Plan for revision: *We will add that the irrigation signal in the SMOS-DISPATCH dataset is a product of the downscaling method that redistribute the 40km mean SMOS SM based on MODIS NDVI and land surface temperature.*

In Section 4, abbreviations for the approaches are introduced. It would be useful if in the former section, a table is created that summarizes all evaluated approaches and defines the abbreviations for readers to reference. This would be particularly helpful because the abbreviations are used in Figure legends.

Reply: *Thanks, indeed a very useful table to keep track of the various approaches*
Plan for revision: *Such a table will be added to the manuscript.*

Table 1: Can a figure be added either to main text or supplementary that shows comparisons of the model and observed SM and ET throughout the period? This can be useful to visualize bias characteristic (e.g., random vs. systematic)

Plan for revision: *Time series will be added to the supplementary material showing the comparison of observed and modeled soil moisture.*

Figure 3: please use different colors to differentiate between NS-SM_bf and RS-SM_bf. Please add the benchmark to the legend. Is there a reason the benchmark is shaded instead of a line (like the predictions)? (Similar sentiments for Figures 4 & 5 aesthetics)

Reply: *Thanks for the comment. For figures 3 and 5 we would argue that having different plotting e.g. shaded for benchmark and line plots for results makes it easier to visualize the similarities and differences. In Figure 4 the goal is to compare the calibration approaches which is why we choose the joint approach as a shaded plot behind the separate approaches which makes it easier to show how the contribution from soil moisture and evapotranspiration varies according to the calibration strategy.*
Plan for revision: *We will add the benchmark to the legend in Figure 3 and find other colors for NS-SM_bf and RS-SM_bf.*

Section 4.2 can benefit from more attention to writing to report results in a more clear and concise manner

Plan for revision: *We will try to rewrite parts of the section to make it more concise.*

Paragraph starting in line 333: Is this bias characteristic model specific to the mHM? If calibration considered other metrics (e.g., NSE) could this error source be reduced?

Reply: *We did not test other hydrological models to see whether this behavior is specific to the mHM model, but the mean and standard derivation of simulated soil moisture will always relate to the parameters related to wilting point and saturated water content. Such parameterizations are found across hydrological models We do not think of this as a model issue, but rather an issue related to the SMOS-DISPATCH datasets mean-variance relationship that is against our expectations, which could relate to uncertainties in the downscaling method. Choosing another calibration metric would not have solved this issue because of the final rescaling between the modeled and satellite mean soil moisture, which was necessary for the model to have a more comparable reaction to the precipitation with respect to the satellite-based soil moisture.*
Plan for revision: *None*

Line 392: comma after “Figure 6” should be removed

Plan for revision: *Will be removed from the manuscript.*

Section 4.3: It would be extremely helpful to the community if this paper summarizes the insights from analyses and comparisons in this section. Namely, a table which summarizes the strengths and weakness (e.g., uncertainty sources) of the approaches, and which approaches are more robust to various uncertainty sources (e.g., precipitation, satellite temporal and spatial resolution, noise, etc.). This table would likely be the primary take-away of the study.

Reply: *Thank you for this comment.*
Plan for revision: *We will add a table that summarizes some of the outcomes of this study.*