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

REVIEWER 2

This study focuses on comparing different methods for estimating irrigation water use. The topic is important for understanding the human impact on the hydrological water cycle, and there are various methods available with their own advantages and disadvantages. The authors conducted a comparison between a baseline model representing rainfed agriculture and a satellite-based model that captures irrigation. The difference between the two models represents the unmodeled process, which in this case is irrigation. To ensure accuracy, the authors calibrated the model using rainfed pixels to remove biases between the model and satellite observations before calculating the difference. Other irrigation estimation methods, such as the water inversion method using satellite soil moisture and evapotranspiration (ET), are also used and discussed. The paper provides a comprehensive comparison of different methods, but some details are skipped, possibly assuming that readers are familiar with previous related papers. One concern raised is the possibility of double-counting water when estimating irrigation using both soil moisture and ET residuals, as these variables are interconnected. Overall, the paper is recommended for acceptance after addressing the mentioned comment and the below detailed comments.

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.*

Major comments

It appears to me that author assumed the reader has already read their previous papers so they did not explain some terminologies or hypotheses that they had in the abstract, for instance in L15-17 it is not clear what are the satellite or the rainfed framework and what they meant by the baseline framework. Later in the introduction at L43-44, the study hypothesis is explained, I think something to this effect can be added to the abstract.

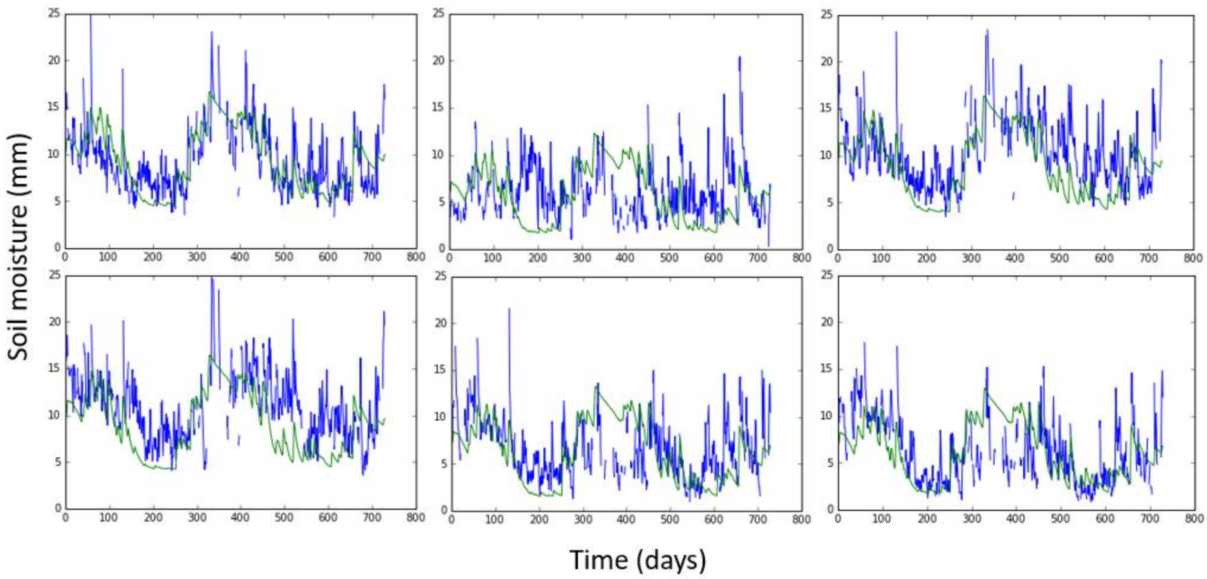
Reply: *Thank you for your comment.*
Plan for revision: *We will try to make this clearer in the manuscript, especially focusing on the definition of key terminologies upfront in the introduction.*

I do not understand lines 236-244 regarding how the water is not counted twice. The ET and soil moisture are interconnected, such that an increase in ET results from an increase in soil moisture. Technically some of the water that enters the soil and increased the soil moisture will later be consumed by the plant with a delay and transpired into the atmosphere. Moreover, it is not necessarily from the rootzone, some studies showed that plants' roots get most of the water from topsoil rather than deeper layers. Thus, the water that is once accounted as residual soil moisture will be later extracted by the plant root and then accounted for twice in the calculations.

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.*

L235-236 & L304-305: If the soil is over-irrigated then the surface soil will become saturated and SM will stay at a constant level. Consequently, it would not be able to reflect both soil moisture storage and ET fluxes change. Please comment on this.

Reply: *It's true that over-irrigated fields would pose an issue and require that drainage and overland flow would also need to be accounted for. However, based on the following plots showing SMOS DISPATCH soil moisture observations for a random set of irrigated pixels, we draw the conclusion that this is not a big concern within our study area. Results from Dari et al. 2020, who used a soil moisture-based inversion framework, also suggested that drainage from irrigation accounted for less than 0.5% of the irrigation within our study area.*



The six plots show the SMOS DISPATCH soil moisture observations (blue) together with the modeled baseline for the RZ_SM_bf approach over two years from 2016-2017.

Plan for revision: *We will mention that a drainage and overland flow term might be necessary to consider if this method were to be used in an area with known over-irrigation.*

Minor comments

Figure 2) There are some positive and negative residuals after calibration for the rainfed cropland that can propagate to the residual estimated over the irrigated pixels, how are these errors treated in your approach?

Reply: *The baseline model does sometimes either over- or underestimates the rainfed soil moisture or evapotranspiration which will affect the irrigation estimates. However, the overall mean error of the baselines is close to zero for all calibration targets within the four approaches, which implies that the error related to the baseline uncertainty will even out over the simulation period.*
Plan for revision: *We will add rainfed time series to the supplementary material from the calibration to illustrate the small bias. The overall SM and ET biases over rainfed land will be quantified as well as the standard deviation of the rainfed residuals.*

L193: Could you add a figure with 4 maps to the manuscript? Firstly add two maps derived from ET and soil moisture temporal stability analysis. Secondly, create an overlap map that combines ET and SM maps. Then compare it with an independent land use map that shows the rainfed and irrigated cropland and report how accurate was the rainfed cropland mapping. This step is crucial as the bias removal process is conducted based on the selected pixels from these maps.

Reply: *Thank you for this comment.*

Plan for revision: *We will add two maps to the supplementary materials showing the results from the ET and soil moisture temporal stability analysis separately. The combined results can be seen in Figure 1 compared with an aerial photo from where there is a clear difference between the green irrigated and beige rainfed cropland.*

L197: By calculating the MAE spatially at each time step, you won't have individual values for each pixel. Instead, there would be only 10 or 14 values for the calibrating parameters after optimization. However, wouldn't it have been more beneficial to have separate values for each pixel by calibrating for all the pixels during the non-irrigated period?

Reply: *Thanks, we did optimize by calculating MAE for all pixels during the two-year calibration period. Indeed it would have been optimal to calibrate the model over both rainfed and irrigated cropland during the non-irrigated period, but we know that some irrigation does occur within the districts throughout the year which is why we decided only to include exclusively rainfed cropland in the calibration.*
Plan for revision: *We will try to better describe how we calculated the MAE and choice of a target area.*

In equation 1, how time is considered in the calculation of BAE are you again averaging MAE for all time steps? If yes, please show that in the equation and also mention this in the text.

Reply: *The mean absolute error (MAE) is calculated on a pixel level and not by using an area average like the irrigation estimates.*
Plan for revision: *We will make this more clear in the manuscript.*

L205-210: I am having trouble understanding the steps described in L205-210. Are you implementing both model calibration through optimizing an objective function and rescaling by adjusting the model mean SM to match the satellite SM mean? Is rescaling necessary or has it already been accounted for in the calibration process?

Reply: *Yes, we are implementing both in the calibration process due to a model/observation issue discussed in lines 333-334.*
Plan for revision: *We will make try to make this more clear in the manuscript.*

L236: replace “it” with water

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

L270: what is the rootzone soil moisture data you used for calibration?, mention it here for RZ_SM_bf

Reply: *The root zone soil moisture is calculated from the SMOS DISPATCH topsoil soil moisture product by using an equation from Albergel et al. 2008 that estimates the amount of entering water into the soil that represents the root zone after two days from entering the soil.*
Plan for revision: *We plan to add a table that summarizes all the models and different inputs for a better overview.*

Figure 4) I find it a little bit confusing to interpret the legend in Figure 4 legend. To enhance the legend clarity I suggest putting the ET and SM of each approach in individual boxes and labeling them with the corresponding approach names. This adjustment would make it easier to comprehend the legend. Additionally, the solid blue line is not explained in the figure caption.

Reply: *Thanks for your suggestion.*
Plan for revision: *We will try to make the legend more clear.*

L315-316: How can we ensure that storage of the water from the previous season is the primary reason for estimating a higher irrigation value compared to the benchmark and not an overestimation of irrigation by the model?

Reply: *We don't have any data to assure that some of the estimated irrigation is water stored from last season. However, in Algerri Balaguer we know that this is common practice and we know that reservoirs exist within all irrigation districts which allow the farmers to distribute the water as needed. If we look at Figure 2, the use of storage water in Algerri Balaguer in April 2017 is clear, but not within any other districts. If this irrigation peak were caused by a large model bias, we would expect a similarly large increase within the surrounding districts.*
Plan for revision: *We will comment further on this in the manuscript.*

L325: Perhaps the reservoir is operating based on a relatively fixed plan, regardless of how much precipitation is received. Have you explored this possibility?

Reply: *Yes, we also think that makes sense that large reservoirs like this work with a relatively fixed schedule based on the reservoir capacity, incoming water, and when irrigation water normally is needed within the districts. We do not have contact with the reservoir managers to get this confirmed.*

Plan for revision: *Add a comment on the possibility of a fixed irrigation schedule to explain benchmark data in relation to the actual irrigation practice.*