Reviewer#2: Evaluation and comments

General Comment:

Irrigation is the largest human intervention in the water cycle, yet it is poorly represented in the land surface (LSM) and hydrological models. One way to account for irrigation is by assimilating the observations that contain the irrigation signal, such as radar backscatter (σ^0) or satellite soil moisture retrieval. One important step prior to the assimilation is removing biases between the model and the observation through calibration. In this study, the Noah MP model is coupled with a backscatter observation model (WCM), and Sentinel-1 σ^0 observations were used to calibrate the model. Furthermore, the impact of activating the irrigation schemes within the Noah MP model, using different backscatter polarization (VV or VH) and cost functions in model calibration, is investigated. I found the study interesting; however, as a calibration study, I expected that the results be more focused on the calibration rather than evaluating the performance of the LSM. I also have some concerns regarding the mixed results obtained regarding the activation of the irrigation module. Please see my comments for detail:

Reply. We would like to thank the reviewer for this general comment and for the interest in the manuscript. We will implement the following specific comments more in depth.

Specific comments:

1. **L175: The size of the Budrio test sites is much smaller than the model resolution (almost 1/200 of the model spatial resolution). I do not think it is a good choice for validation purposes.**

Reply. Thanks for this comment. We agree that this is a crucial point, especially when modelling human activities such as irrigation. One of the most critical aspects of irrigation validation is given by the lack of irrigation benchmark data (Foster et al. 2020). In this specific case, we decided to not exclude the Budrio test site considering the reliability of the data over the fields. The second aspect is that we realized evaluations at different scales: 1) regional (the entire study area); 2) small-district (Faenza test site); and 3) plot scale (Budrio fields). Indeed, while the Budrio test site is composed by plots of about 0.4 hectares, the analysis over the Faenza test site (see Figure 10) refers to an area of 270 Ha which is comparable with the model estimates. Finally, it has to be noted that we have selected an intensively irrigated area. Maps such as the Global Rainfed Irrigation Paddy Areas (GRIPC; Salmon et al., 2015) confirm that the Po river valley is almost entirely irrigated with a percentage of about 100% over the 0.01° LIS grid. This reduces the risk of finding non-irrigated fields within the 1km LIS grid. We are aware that there are limitations in our approach but we think that 0.01° spatial resolution is a good compromise between analysis on a regional, small-district and plot scale. We plan to add a small section in the revised manuscript to better clarify this point.

2. **L300-305: It is not indicated which vegetation indicator (VWC, NDVI, or LAI) is finally chosen to represent V1 and V2 in equation 2 and 3. Moreover, it is not clear whether it is assumed that V1=V2 and a unique descriptor is used for both of them or not? According to the rest of the paper, I suppose that LAI is the chosen vegetation indicator, but I think this should be explicitly mentioned here.**

Reply. Thanks for this comment. We agree with the reviewer. This point is not explicitly mentioned. We assumed V1=V2, represented by the dynamically simulated LAI vegetation descriptor. We will add this explanation in the text of the revised manuscript.

3. Section 2.4: As mentioned in L95, assimilating the SSM or VWC retrieval instead of MW brightness temperature or σ0 can be problematic due to inconsistent ancillary data used in their production, and $σ^0$ is a better choice for assimilation. However, as is explained in section 2.4, assimilation of $σ^0$ requires the NOAH MP model to be coupled with a WCM model to simulate $σ^0$. In turn, the WCM model has many empirical parameters and simplifying assumptions such as not accounting for scattering interactions between ground and vegetation and assuming a linear relation between soil $σ^0$ and the SSM that can increase the uncertainty of the estimated $σ^0$. Given this, can the authors clarify why assimilating $σ^0$ is a better choice relative to the assimilation of SSM and VWC?

Reply. Thanks for this comment. A valuable reply to this comment is contained in previous papers we have cited in the introduction section at lines 94-101. As reported in De Lannoy et al. (2016), a critical aspect in directly assimilating SSM retrievals is that potentially inconsistent ancillary data are used in the assimilation system and in the retrieval algorithm that generates SSM observations. Furthermore, active MW retrievals typically use change detection methods (Wagner et al., 2013; Bauer-Marshallinger et al., 2018) which lack land-specific information. This means that the 'error management' within the data assimilation system is theoretically more transparent when assimilating backscatter observations. Using microwave retrievals allows us to have consistent parameters between the LSM and the radiative transfer model (in our case the WCM) and to avoid cross-correlated errors.

4. L356: Another interesting comparison would be comparing the performance of a third calibration approach which is, deactivating the irrigation scheme and calibrating the model only during the non-irrigated season, with the current approach (activating the irrigation scheme and calibrating for the entire period) during the non-irrigated season. This would also be an interesting comparison for the future study in which $σ^0$ will be assimilated to see whether calibrating during the irrigation season with the activated irrigation module is beneficial for the ultimate goal of irrigation quantification or not.

Reply. Thanks for this comment. We tested the additional calibration option of removing the irrigated periods (i.e., summer) when the irrigation scheme is deactivated. We found a not reliable saturation of the vegetation parameters (especially the A parameter) through the upper boundary over the cropland areas with wrong simulations during the summer, affected by very high summer peaks in the backscatter signal (especially affecting the backscatter VV polarization). In this specific case, the higher contribution to the signal is given by the SSM which is higher during the winter. On the other hand, the backscatter from the vegetation cannot provide a large contribution as the LAI signal is low during the winter. This prompts the vegetation parameters to be pushed towards their upper boundaries. An example of parameters maps for the KGE-VV Natural experiment, removing the irrigation period, is shown in Figure R2 below. We also added an example time series plot (Figure R2e) where we compare the Sentinel-1 and the WCM signals over a pixels in the cropland area where the A reaches a value of 0.4 [-] (upper boundary). The blue time series, representing the WCM signal, clearly shows the anomalous peaks during the summer period.

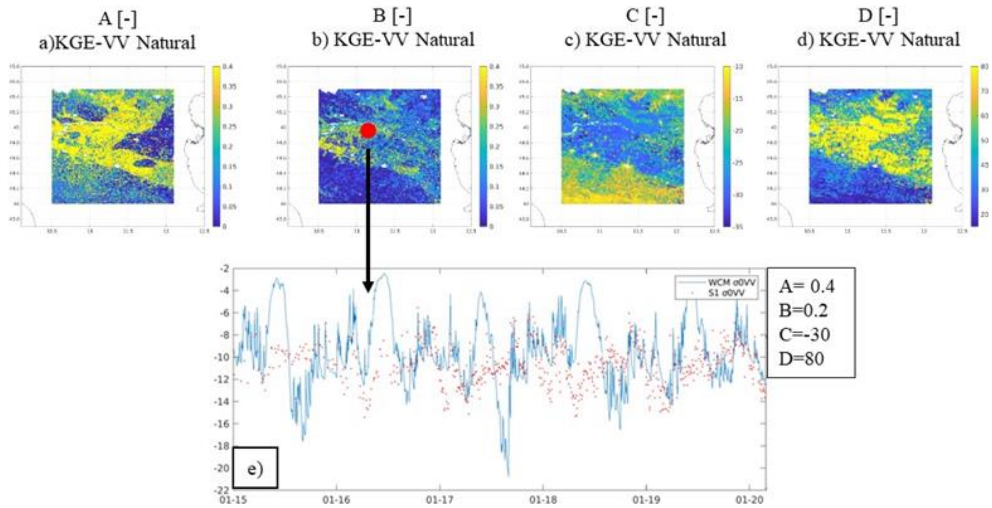

Figure R2. Maps of: a) A parameter; b) B parameter; c) C parameter; d) D parameter for the *KGE-VV Natural* calibration experiment, realized removing summer irrigation period. e) Time series of simulated and observed (Sentinel-1) backscatter for an example grid cell.

We also observed this phenomenon in the Natural experiment considering the entire study period (summer+winter) but at a smaller scale and for an opposite reason as discussed at lines 480-486.

Thanks for the suggestion, we will briefly include this thought and our findings in the discussion of the revised manuscript.

5. L405-416: I am not convinced that the improvement in the simulation of SSM by Noah MP is due to the activation of the irrigation module for two reasons:

- **In most of the previous studies, it is shown that coarse-scale MW products are not able to detect irrigation signals at the plot scale (e.g., Brocca et al., 2019, Zaussinger et al., 2018, Dari et al., 2020) unless there is intensive flood irrigation over a large area such as California central valley in which fields are flooded, and the water level is sustained throughout the irrigation season (Lawston et al., 2017). Moreover, to have a fair comparison, other metrics such as RMSE or bias should also be reported alongside the Pearson correlation (R).**
- **The deterioration in LAI simulation when irrigation is activated makes more sense to me as the spatial resolution of the Proba-V LAI product is 1km, and the possibility of detecting the irrigation signal is higher relative to the coarse-scale SSM products.**

Please comment on this.

Reply. Thanks for this comment. Po river Valley is one of the largest and most intense irrigated areas in Europe so it is very likely that the contribution of the irrigation in the coarse-scale SSM is visible especially at biweekly temporal scale (i.e., the majority of the land within the satellite footprint is irrigated).

On the other hand, for the second point, we agree with the Reviewer. The higher spatial resolution of the Proba-V results in finer spatial details about crop state due to plot-specific agricultural practices, the unknown yearly variability of the crop types and the impact of the meteorological conditions in the stakeholders decision process as we already explained at lines 568-580 of the discussion section which are difficult to represent when irrigation is activated (for the simplified model parameterization of the irrigation process). Furthermore, temporal dynamics of LAI are clearly more sensitive to root zone soil moisture which might be more difficult to simulate than SSM during the irrigation season due to very likely larger impact of the soil texture and

transpiration processes along with the high frequency of the wetting and drying caused by irrigation events.

Additional discussion will be provided on these aspects in the revised version of the manuscript where also other metrics will be used as suggested by the reviewer.

6. L435-446: I think the part of this significant difference between simulated and observed irrigation and missing the irrigation events is related to the spatial mismatch between the model and the test site, as it is mentioned in the first comment. Please comment on this.

Reply. Thanks for this comment. The reviewer can refer to the reply to comment (1)

References

- Bauer-Marschallinger, B., Freeman, V., Cao, S., Paulik, C., Schaufler, S., Stachl, T., Modanesi, S., Massari, C., Ciabatta, L., Brocca, L. and Wagner, W., (2018a). Toward Global Soil Moisture Monitoring With Sentinel-1: Harnessing Assets and Overcoming Obstacles. IEEE
- De Lannoy, G.J.M., Reichle, R.H. Assimilation of SMOS brightness temperatures or soil moisture retrievals into a land surface model Hydrol. Earth Syst. Sci., doi:10.5194/hess-2016-414, 2016
- Foster, T.; Mieno, T.; Brozovic, N. Satellite-based monitoring of irrigation water use: assessing measurement errors and their implications for agricultural water management policy. Water Resour. Res. 2020, 56, 10.1029/2020WR028378
- Salmon, J. M., Friedl, M. A., Frolking, S., Wisser, D.: Global rain-fed, irrigated, and paddy croplands: A new high resolution map derived from remote sensing, crop inventories and climate data, International Journal of Applied Earth Observation and Geoinformation, 38, 321–334, doi:/10.1016/j.jag.2015.01.014 , 2015.
- Wagner, W., Hahn, S., Kidd, R., Melzer, T., Bartalis, Z., Hasenauer, S., Figa-Saldaña, J., de Rosnay, P., Jann, A., Schneider, S., Komma, J., Kubu, G., Brugger, K., Aubrecht, C., Züger, J., Gangkofner, U., Kienberger, S., Brocca, L., Wang, Y., Blöschl, G., Eitzinger, J., and Steinnocher, K.: The ASCAT soil moisture product: A review of its specifications, validation results, and emerging applications, Meteorol. Z., 22, 5–33, 2013