Dear Editor,

We appreciate the review of our manuscript and we have thoroughly addressed all comments by the reviewers and editor. This involves major textual clarifications, an update to figure 1 and a new appendix to explain model differences in ET calculation, but the scientific findings have not changed. The responses to the reviewers are already published. We provide point-by-point responses to the editor comments (in boldface) below. Our responses are in normal font and parts from the revised manuscript are in italic font.

1. Dear Authors, Your original manuscript has been evaluated by two reviewers who provided slightly different ratings. Rev.#1 was critical of the scientific quality and significance of this study and recommended it should be rejected altogether. Rev.#2 was more positive but ultimately suggested major revisions. Helpful comments from one discussant (Nima Zafarmomen) were also received by the paper, although no response was provided.

We now also responded to the community comment from Nima Zafarmomen (shown in bold grey) in the interactive discussion:

> **The paper does an excellent job comparing AquaCrop and Noah-MP, two models with distinct goals and methodologies, to evaluate their performance in estimating irrigation in the Po Valley. This side-by-side analysis provides valuable insights into the advantages and limitations of crop models versus land surface models when applied to complex irrigation systems.**
>
> **1) How might data assimilation of satellite soil moisture or vegetation index data affect the model's ability to capture interannual variability in irrigation accurately? Would this approach potentially reduce some of the overestimations observed?**
>
> **2) Given the paper's findings on the variability of irrigation rates and timing, do the authors see data assimilation as a viable approach for operational large-scale irrigation modeling, or do they foresee challenges in scalability?**
>
> **3) Noah-MP includes canopy interception and runoff losses, whereas AquaCrop does not. Could the authors discuss how these differing approaches to irrigation losses might affect their model outputs, especially in the context of large-scale regional studies?**
>
> **4) I highly recommend the authors consider citing recent work on data assimilation in hydrological modeling for irrigation, specifically studies that integrate satellite-based vegetation indices to improve model accuracy. For example, studies like 'Assimilation of Sentinel-based Leaf Area Index for Modeling Surface-Groundwater Interactions in Irrigation Districts'and 'Multivariate Assimilation of Satellite-based Leaf Area Index and Ground-based River Streamflow for Hydrological Modeling of Irrigated Watersheds using SWAT+' showcase how data assimilation can enhance the irrigation districts.**
>
> We would like to thank Nina Zafarmomen for posting this community comment. We answer the four questions below:
>
> First, a general response relating to the comments 1, 2, and 4: the objective of the study was to show the difference between the models as they are and are currently used, not to

perform a data assimilation (DA) analysis. DA is certainly an interesting approach to improve the results, we work on it ourselves and it is mentioned in the original (and revised; L719-723) manuscript, but it is beyond the scope of this paper. However, we added the requested citation in our paper and provide specific answers to the comments as follows:

1. We fully agree that vegetation DA would improve the vegetation estimates (along with other modeled variables) in both models, if the DA is correctly setup, meaning that the observations should be unbiased compared to model estimates (Scherrer et al., 2023). Additionally, we want to emphasize that the CGLS DMP product used as a reference in this paper should not be taken as the absolute truth and may be biased compared to reality. The limitations of the validation datasets are now further discussed in the manuscript.
2. Filtering (or sequential) DA in irrigated areas has shown limitations in improving irrigation estimates when irrigation is explicitly represented in LSMs, as e.g. demonstrated in Busschaert et al. (2024). Most importantly, the spatial and temporal scale remain an issue, as irrigated fields are typically smaller than the resolution of this study (0.01°) in Europe and satellite DA thus requires high resolution vegetation and soil moisture observations. Furthermore, irrigation can be applied one morning and satellite observations of soil moisture may only see a remnant a (few) day(s) later at overpass time.
3. The manuscript thoroughly discusses the differences in irrigation losses between the two models in Sections 3.1.4 and 4.1. The canopy interception losses are indirectly considered in AquaCrop as stated in L214-216 from the revised manuscript: "*AquaCrop does not consider canopy evaporation (or leaf interception loss) explicitly, but it is indirectly included because the intercepted water is assumed to infiltrate into the surface soil, from where it can be lost by evaporation*". The runoff losses are indeed different, but by considering typical application losses, the average annual irrigation rates from both models are aligned.
4. We thank you for the references: even though slightly out of scope, we added Igder et al. (2022) on L722.

References:

Scherrer, S., De Lannoy, G., Heyvaert, Z., Bechtold, M., Albergel, C., El-Madany, T. S., and Dorigo, W.: Bias-blind and bias-aware assimilation of leaf area index into the Noah-MP land surface model over Europe, Hydrol. Earth Syst. Sci., 27, 4087–4114, https://doi.org/10.5194/hess-27-4087-2023, 2023.

2. In my opinion, the authors' responses to the criticisms raised by the two reviewers during the discussion were somewhat weak. Nevertheless, there is considerable room for further improvements so as to make the revised paper appreciated by the HESS readership and a solid contribution to a journal of high standing such as HESS.

The edits suggested in the responses to R#1 and R#2 were slightly updated, in light of the Editor's comments. In addition to those edits, we now make it very clear that this research is bridging between two scientific communities, each with their own modeling frameworks. We further edited the title, the abstract, the discussion, and the conclusions. The specific edits are described in the responses to the comments below.

3. First, and as suggested by both reviewers, the paper is lacking a clear statement about the goals of the study and the various steps implemented by the authors to achieve these goals. In their responses, the authors have acknowledged that point, but I think the question needs to be treated more in-depth. The research questions are not clear to me and, at the closure of the introduction section, there is only information about the various parts of the paper. The authors wish to add the following sentence at L37-40: "*The irrigation modeling is not based on agricultural practices, but has been included to better simulate the water balance, thereby improving all estimates of water, energy and carbon fluxes*". I hope I did not misunderstand, but how could it be otherwise? If, in a certain agricultural area, there are not only precipitation and evapotranspiration fluxes but also irrigation events, the latter must be modeled to accurately assess the water balance of the soil-vegetation-atmosphere system.

We have now further clarified the goals of the study upfront. First, we adapted the title of the manuscript to "*On the gap between crop and land surface models: comparing irrigation and other land surface estimates from AquaCrop and Noah-MP over the Po Valley*".

Second, the following lines were added at the beginning of the abstract to better state the goals of the study:

Land surface and crop models both simulate irrigation, but they differ in their approaches, primarily because they were originally developed for distinct purposes and scales. Through an example case study in a highly irrigated region, this research helps to better understand the differences between these models and the complexity of irrigation modeling. More specifically, irrigation was estimated over the Po Valley (Italy) at a 1-km^2 spatial resolution using (i) a crop model, AquaCrop, and (ii) a land surface model, Noah-MP. [...]

Third, the end of the introduction section is updated as follows:

L71-84: *[...] The objective of this study is to evaluate irrigation estimates and other variables from a crop model and an LSM at the regional (basin) and pixel scale (field-based evaluation) [...] Both AquaCrop and Noah-MP have been used to estimate irrigation in previous studies in their respective scientific communities [...]*

The requirements for better observations, which are now discussed in the revised manuscript, are also stated at the end of the abstract:

The results of this study highlight the complexity of irrigation modeling due to its anthropogenic nature, but also show the need for better observations to guide model estimates: reference irrigation data are sparse and satellite retrievals under irrigated conditions are quite uncertain.

Regarding cited L37-40 of the revised manuscript, we think that the sentence is valuable when put into context. The sentence is now updated and put back in context below:

L33-43: *Originally, LSMs were mainly concerned with the calculation of surface energy and water fluxes, but these models have grown in complexity, with modeling advances for e.g. vegetation, snow, soil moisture, and more recently, the implementation of crop and irrigation modeling (Fisher et al., 2020). The LSMs were developed for coarse spatial resolutions (0.5-2°) and have been gradually used at finer resolutions (Fisher et al., 2020), but they most often do not resolve individual fields. The irrigation modeling thus does not aim to exactly reproduce agricultural* *practices, but has been included in LSMs to better simulate the water balance, thereby improving all estimates of water, energy and carbon fluxes. Despite its importance, irrigation remains often unmodeled or oversimplified (McDermid et al., 2023).*

4. The authors state in the highlights that "the variability of simulated irrigation differs from satellite observations". However, I could not find an adequate explanation for this mismatch in the paper, as well as in their responses to the reviewers. This comment is closely related to some criticisms raised, both directly and indirectly, by both reviewers regarding the lack of analysis of the epistemic uncertainties associated with the two tested models (i.e., AquaCrop and Noha-MP). I believe that this analysis should be a key component of the revised paper since it will help understand the different outcomes of the two tested models.

This sentence is not present in the manuscript submitted to HESS. We are thus afraid an outdated and non-reviewed submission to the Journal of Hydrology might have been consulted. We moved our submission from JoH to HESS because of a better fit in scope, and had removed the highlights, because these are not considered for publication in HESS. Nevertheless, the difference in the variability of irrigation from the models (AquaCrop and Noah-MP) and the satellite-based retrievals from Dari et al. (2023) is described and discussed in Section 3.1.2, when describing Figure 5:

L443-458: *To better understand the models' skill at different time scales, the irrigation amounts were aggregated (sum) over different time intervals: from weekly (original resolution of the retrievals) to yearly (based on the summer months only JJA). The temporal R is then calculated between the datasets (AquaCrop, Noah-MP and ref) for each pixel and the results are presented as spatial boxplots in Figure 5. When aggregating over longer time intervals, both models tend to show a higher correlation (grey boxes), meaning that they capture the same high- and low-irrigation years. However, when computing the correlation with the reference data from Dari et al. (2023), R tends to deteriorate with increasing temporal aggregation with negative R values in several locations. If the reference data have the right temporal variability, this would prove that the models are not capturing the actual interannual variability (or the absence of interannual variability) as they mainly depend on the meteorological forcings and are not restricted by e.g. water availability. However, it will be shown that both AquaCrop and Noah-MP tend to better capture the interannual variability in irrigation when compared to in situ data (later discussed in Section 3.2), which suggests possible inaccuracies in the reference data.*

5. The uncertainty analysis has to be complemented by a sensitivity analysis to evaluate how, and to what extent, the uncertainty in the input variables is transferred to the outputs. For example, and if I understood well, topsoil soil moisture values were not measured; consequently, the reader is unaware of the uncertainty associated with the downscaling of SMAP data and how this uncertainty is transferred to the model outputs.

Please also see our response to R#2.

First, based on our experience and literature (Koster et al., 2009; Crow et al., 2023), <u>models are complex and therefore developed with their own set of parameters</u>. As explained in the original manuscript, we attempted to use the same soil hydraulic parameters in both models, but this resulted in even larger discrepancies in ET from both models. We now make this point even stronger and supplement it further with literature:

L672-678: *Additional simulations were performed (not shown) using the more realistic soil parameters for field-based applications as used in AquaCrop, but the results of the Noah-MP and AquaCrop models then diverged even further, with a drastic increase in ET for Noah-MP. This further underscores that each model is developed with its own set of parameters, highlighting that soil moisture is a model-dependent quantity (Koster et al., 2009), and that each model has its own coupling mechanisms between soil moisture and fluxes of ET, runoff (Crow et al., 2023) and irrigation.*

Second, from our experience with both models, we believe that <u>conducting a sensitivity analysis would result in a standalone study in itself</u>. For example, a sensitivity (calibration) study on only the impact of one parameter (the irrigation threshold) in Noah-MP is part of another manuscript that is currently under review:

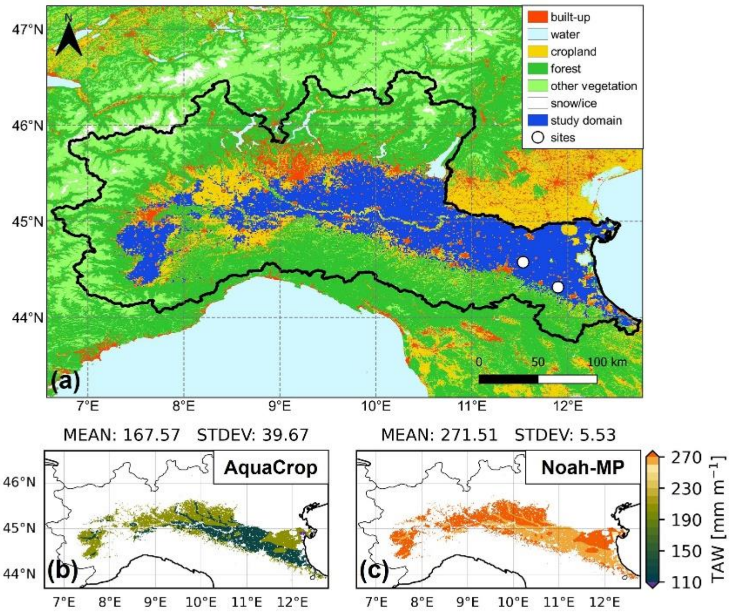
Modanesi, S., Busschaert, L., De Lannoy, G., De Santis, D., Natali, M., Dari, J., Quintana-Seguí, P., Castelli, M., Massimo Grasso, F., and Massari, M. 2025. Accounting for scaling effects on irrigation optimization within a land surface model using satellite observations: impacts on water and carbon cycle dynamics. In review for Journal of Hydrometeorology.

Lastly, performing a sensitivity analysis on the models by varying their parameters <u>would not provide additional insight into the uncertainties associated with the downscaled SMAP retrievals</u>. These satellite-based products inherently carry uncertainties, which are typically quantified through comparisons with in situ measurements. We have acknowledged these uncertainties in several parts of the revised manuscript, as also noted in our responses to R#1 (comment 2) and R#2 (comment 4). For these reasons, a sensitivity analysis cannot meaningfully contribute to or strengthen the conclusions of this paper.

6. A weakness of the work seems the characterization of the soil hydraulic properties that govern water transfer in the soil module of both models. The (very) simplistic concept of soil moisture at the condition of field capacity (FC) is the basis of AquaCrop, which is a bucket-type model. If I am correct, there is no information on how the FC values are determined in the distinct zones of the study area. The literature on the matter demonstrates that the way FC is determined can be a key issue for obtaining reliable results by a bucket-type hydrological model, especially when one would compare the results of a bucketing-type model (such as AquaCrop) with a Richards-based model (such as Noah-MP). On the other hand, Noah-MP requires information on the soil-water retention and hydraulic conductivity functions, but I only found the reference to the papers by Cosby et al.(1984) and by Chen and Dudhia (2001). This is not sufficient, and specifications are required for the soils of the study area. Whereas both models may be less sensitive to the soil moisture values at wilting point (WP), without a clear correspondence between the FC values (or, the soil hydraulic properties) and the actual soil conditions, the results of this study appear to be nothing more than a modeling exercise from which it is also difficult to draw conclusions. This aspect makes the whole paper rather weak for publication in HESS. I am sure the authors will resolve that question when revising the paper.

We thank the Editor for this constructive comment. The soil texture classes were extracted from the Harmonized World Soil Database v2.1 and were identical for both models. The soil hydraulic parameters of each soil texture indeed differ between models because of the reasons outlined in the response to the previous comment.

To help the reader understand the processes and to visualize the difference in SHPs, Figure 1 was divided into Figure 1a (original Figure 1) and Figure 1b-c presenting the total available water (TAW) over the domain, a key parameter for the irrigation modeling in both models.



(a) Map of study domain showing a combination of the different CGLS land cover classes and the GRIPC irrigation classification. The rainfed cropland class and paddy rice are shown in yellow, different forest types are aggregated into one forest class in dark green, and the remaining vegetation classes are in light green. The dark blue areas correspond to irrigated cropland (excluding paddies) of interest to this study. The Budrio and Faenza sites are marked with white dots. (b) and (c): maps of the TAW [mm m^{-1}] derived from the SHPs of each model.

The figure is introduced in the methodology:

L166-168: *Figure 1b and c show the TAW over the domain for each model (i.e. set of SHPs). Across the different soil classes present in the domain, the TAW ranges are 80-200 and 255-276 mm m^{-1} for AquaCrop and Noah-MP, respectively.*

7. The issue of surface runoff should deserve adequate clarification in the revised paper. The original paper reports several comments about surface runoff and how this variable is obtained by the two models. However, given the absence of topographical data in the study area, the question remains unresolved regarding whether full saturation of the uppermost soil layer results in runoff or merely leads to ponding of water on the land surface. The authors should acknowledge that their models, particularly AquaCrop, adopt the simplification of a uniform soil profile up to the selected maximum rooting depth of 1 meter. Consequently, apart from the greater or lesser presence of clayey soils in specific zones of the study area, it is the actual presence of soil layering that also concurs to govern the occurrence of surface runoff or ponding water. The authors should discuss the issues of surface runoff and ponding water in light of these comments.

These are indeed two important aspects to clarify. Those points are now discussed:

L699-707: *In both models and, more in general within one-dimensional modeling frameworks, the water lost through runoff, as detailed in the regional water balance analysis (Section 3.1.4), is* *removed from the pixel, resulting in a loss from the system, even though runoff plays an important role in local and regional water management. Nevertheless, since irrigation primarily occurs during the summer period, surface runoff has a milder impact on the results. Another limitation of both models is the assumption of a uniform soil texture profile, which limits the accuracy of the TAW content, a critical parameter in irrigation modeling. Although a configuration including both topsoil and subsoil layers for a heterogeneous soil profile is desirable, it is not yet used within NASA's LIS. [...]*

8. Allowing for the reviewers' and discussant reviews, as well as my editor report, the paper requires substantial revisions. The authors are invited to submit a revised version together with detailed point-by-point replies to all the comments received so far. Should you disagree with a specific comment, please explain the reason clearly. The revised paper will require an additional round of reviewers' evaluations, and there is no guarantee that it will be accepted for final publication in HESS. If the authors think that the required changes will take longer than the time allocated for submitting the revision, it is recommended to withdraw the paper but I encourage to resubmit it. In this case, and if the authors agree, I am willing and very happy to handle it again as editor in charge to facilitate some steps in the review process.

We have now addressed all comments in detail. To clarify the main results of this paper, we re-organized Section 4 now called *Overview, limitations, and future perspectives,* into (4.1) *Overview of the modeling results* listing the main findings and discussions from Section 3, (4.2) *Model limitations and potential improvements*, (4.3) *Uncertainty and scarcity of in situ irrigation data*, and (4.4) *Uncertainty in satellite-based retrievals over irrigated areas*. Section 4 therefore highlights the main conclusions on irrigation modeling with either crop models or LSMs by leveraging the results of the modeling case study.

A revised version of the manuscript is uploaded, including all the revisions mentioned in the responses to the reviewers and to the Editor. We are confident that the manuscript has improved, and we understand the request for additional reviewers: we will be glad to receive another round of constructive reviews.