

We thank the four reviewers for their thorough and constructive feedback. We have carefully considered all comments and have revised the manuscript accordingly. We improved methodological descriptions, refined several figures and captions, and strengthened the discussion of uncertainties and implications. In particular, we further clarified the scope and objectives of the study. This manuscript is intentionally interdisciplinary and aims to bridge crop modeling and land surface modeling. Its central objective is to highlight how differences in model structure and process representation translate into different irrigation estimates. In this context, the Po Valley is used as an illustrative study region to examine these modeling differences rather than as a case for crop-specific or regionally optimized simulations. We left model-intrinsic structures and their parameters untouched (generic crop representation, soil hydraulic lookups, ...) because harmonizing parameters to remove model differences would undermine the scientific purpose of the comparison.

We earlier provided a general response to the reviews to explain how we harmonize some differing viewpoints among them, and now provide a point-by-point response. Reviewer's comments are in normal black font, our *responses are in blue* and *parts quoted from the revised manuscript are in italic font*. Newly added or modified text is underlined or shown with strikethrough for clarity.

Reviewer 1

General Assessment

The manuscript is well written, logically structured, and makes an important contribution by comparing AquaCrop (a crop model) and Noah-MP (a land surface model, LSM). This work effectively bridges the crop water and land surface modeling communities, providing valuable insights into irrigation, soil moisture, vegetation, and ET representation at both basin and field scales. The use of multiple observational and satellite products adds robustness. However, some methodological and interpretational aspects require clarification and further depth.

We thank the reviewer for their time and constructive comments. We provide a point-by-point response to the comments below.

Major Comments

1. Section 4.4 – Uncertainties in satellite-based retrievals: The treatment of uncertainties is currently too brief. Methodological uncertainties should be discussed in more detail:

- **Soil moisture representation:** The use of surface soil moisture (SSM) is problematic for irrigation studies, since what is required is modeling of root-zone soil moisture (RZSM) and the management allowable depletion (MA) in irrigation modules to trigger irrigation. This mismatch and the limitation of SSM in representing RZSM should be explicitly acknowledged.
- **Downscaling errors:** Both SenET-derived ET and downscaled SMAP SSM rely on downscaling methods, which propagate errors into the reference datasets and thus into the model evaluation. It may be more consistent to use an ET reference dataset at a coarser resolution that avoids such downscaling errors.
- **Implications:** These uncertainties influence not only irrigation assessment but also the reliability of the ET and SSM benchmarks used in the study. A deeper discussion would strengthen the credibility of the conclusions.

The section on “Uncertainty in satellite-based retrievals over irrigated areas” already highlighted the limitations of satellite-based retrievals in irrigated regions and their implications for this study. We now additionally emphasize the role of downscaling and representativeness errors.

L608-615: The majority of products derived from remote sensing show stronger limitations in the context of irrigated areas than rainfed croplands. First, the native spatial resolution of the remote sensing products (relatively coarse compared to individual fields) leads to substantial heterogeneity within a pixel, both due to mixed land cover and variability within croplands, which complicates the detection of irrigation signals in soil

moisture or vegetation observations (Ozdogan et al., 2010). Downscaled products, such as the SMAP 1-km and 100-m SenET products used in this study, show potential but inevitably carry downscaling errors (Fang et al., 2022) that can propagate into the evaluation of model estimates. Also, for SSM retrievals in particular, a representation mismatch exists between the retrieved SSM and the actual root-zone soil moisture that is most relevant for irrigation studies (Laluet et al., 2024).

We also acknowledge the risk of averaging a downscaled product and the discrepancies between existing ET datasets.

L299-306: The ET is derived using the two-sourced Energy Balance model that downscales the original 1-km Sentinel-3 land surface temperature using Sentinel-2 surface reflectance. For consistency with the model grid, the data were reprojected by spatial averaging. As a result, the ET estimates remain primarily constrained by the original 1-km Sentinel- 3 land surface temperature; however, the re-aggregation may introduce some uncertainty. Differences in spatial and temporal resolution among ET products can influence evaluation results. Nevertheless, substantial inconsistencies among existing ET products have been documented in previous studies (De Lannoy et al., 2024; Modanesi et al., 2025), and no definitive reference dataset for ET currently exists. SenET was therefore selected for its potential suitability for irrigation management applications (Bartkowiak et al., 2024; Chintala et al., 2022; Spiliotopoulos et al., 2023).

2. Temporal aggregation of evaluation metrics: The aggregation of daily values into 15-day means is understandable to reduce the impact of uncertain irrigation timing, but it introduces serious limitations:

- **Loss of dynamics:** Short-term variability in ET and SSM, especially during crop water stress, is smoothed out. This may reduce the evaluation to potential ET comparisons.
- **Bias masking:** Aggregation may artificially improve correlations while masking systematic model deficiencies in simulating fast drying, irrigation events, or recovery after stress.

The performance should be quantified at multiple aggregation scales (e.g., daily, weekly, bi-weekly), which would also be more coherent with irrigation comparison at different temporal scales. Discuss implications for agricultural applications, where short-term stress is critical.

We agree that event scale metrics are important for irrigation scheduling but daily timing accuracy is neither expected from these mechanistic models nor relevant to the overarching goals of the work.

Aggregation smooths short-term variability in ET and soil moisture, but evaluating modeling processes, such as fast drying or stress recovery would require accurate event timing, which is not available at the spatial scale of this study.

Aggregation does not mask systematic biases, as biases scale proportionally with the aggregation period, as shown in the field-based evaluation (Fig. 9). The daily bias can therefore be inferred from the metrics computed at aggregated temporal scales. We now make this explicit in the manuscript for clarity:

L337-339: By definition, the bias scales linearly with temporal aggregation (for irrigation, SSM, ET) and can therefore be consistently related across aggregation levels.

3. Guidance for the irrigation modeling community: Given the comparative scope, the study should conclude with more explicit guidance:

- How to reconcile crop models' explicit treatment of management practices with generalized LSM schemes.
- Implications of shared assumptions (e.g., optimal irrigation rules at ~50% TAW) for large-scale irrigation simulations. Moreover, in the case of sprinkler irrigation, the adoption of such low MA thresholds, essentially at the onset of water stress, appears unjustified, since sprinkler systems enable the frequent application of small water amounts, as evidenced in Fig. 8d.
- Pathways to integrate crop model process detail into LSMs.

We thank the reviewer to bring up these points making a concrete point that the LSM and crop modeling communities can help each other. We have added the following text to the "Model limitations, potential improvements and future perspectives" section:

L561-564: To better represent irrigation (and crop processes), the LSM community is transitioning toward the use of dedicated submodels, such as Noah-MP-Crop (Liu et al., 2016a), in which crop-specific processes and parameters are explicitly represented. This shift inherently moves LSMs toward finer spatial scales (or tiling), helping to bridge more detailed crop model processes with the broader LSM framework.

L575-578: Compared to LSMs, the role of crop models is distinct: they do not aim to simulate the full suite of land surface processes. Yet advances in this research field are increasingly being incorporated into coarser-scale LSMs, provided they are generalizable enough, which is something crop models can further refine by learning from the LSM community.

L579-580: Shared assumptions of both models for large-scale applications (e.g. growing season, irrigation practice, threshold) introduce constraints that might affect model realism.

In the “Conclusions”, we also highlight the contributions from both modeling communities:

L648-650: Although their original roles are distinct, both communities can learn from each other: LSMs increasingly take up processes from crop models, while the crop modeling community can in turn draw insights from the generalization strategies used in LSMs.

Minor comments.

Model inconsistency in vegetation vs ET (Fig. 6): The observation that AquaCrop overestimates vegetation (positive bias in Fig. 6k) while underestimating ET (Fig. 6l) is surprising, especially in quasi non-stressed conditions (as suggested by the MA threshold and Fig. 8c, where MA remains above 50% of TAW). Although possible causes are discussed in the article, this inconsistency deserves a more in-depth analysis.

We thank the reviewer for raising this point. We added a paragraph to further discuss potential causes:

L434-440: For all three satellite retrievals, it is important to note that, in addition to potential retrieval biases, the 0.01° spatial resolution implies that each pixel represents a mixture of land covers and crop types. This sub-pixel heterogeneity can contribute to apparent inconsistencies in the evaluation. In particular, the resampled DMP and ET may include signals from non-cropland vegetation, whereas both models simulate only croplands, potentially leading to a bias. In addition, mismatches in crop phenology may play a role. The generic summer crop used in AquaCrop differs from cropping systems that include winter crops such as wheat, for which ET typically increases earlier in the season. These differences in land cover composition and phenological timing can therefore partly explain the contrasting biases observed in vegetation and ET.

L47-49: Right, but irrigation and other agricultural practices significantly alter water, energy, and carbon fluxes. LSMs need to better represent these processes.

We rephrased this part to clarify this point:

L51-55’: Consequently, irrigation modeling in LSMs does not aim to reproduce detailed agricultural management practices. Instead, it is included to represent the dominant effects of irrigation on the land surface water balance, which is essential for improving simulations of water, energy, and carbon fluxes. Despite its importance and strong influence on these coupled processes, irrigation remains frequently unmodeled or treated in an oversimplified manner (McDermid et al., 2023).

L64: The statement ‘irrigation is not an addition to the model to better represent the water balance’ is unclear. In fact, in crop water models irrigation explicitly supplies water to the soil water storage, thereby improving the representation of water balance components (e.g., RZSM, ET) (Olivera-Guerra et al., 2018).”

We agree that the sentence was ambiguous. It has now been rephrased to the following:

L69-70: In addition, irrigation is an intrinsic element of cropland modeling that represents management decisions, rather than a post hoc addition to improve the simulated water balance.

L66: Beyond their use in estimating net irrigation requirements (i.e., crop water needs), crop models have also been applied to estimate actual irrigation water use under a range of regimes, from deficit to excess irrigation, making them highly relevant for large-scale irrigation assessments (Olivera-Guerra et al., 2023; Laluet et al., 2024).

This is now added to the revised manuscript:

L72-74: Beyond estimating crop water needs, crop models have also been applied to simulate actual irrigation water use under diverse management regimes, ranging from deficit to excess irrigation, which makes them particularly relevant for large-scale irrigation assessments (Olivera-Guerra et al., 2023; Laluet et al., 2024).

L341-342: If runoff is not generated after irrigation in AquaCrop, percolation losses should be included explicitly as residuals in the water balance.

As noted in the model description, the irrigation infiltration is only limited by K_{sat} and is non-limiting over the domain considered in this study. The simplified soil water model (bucket model) of AquaCrop only generates percolation losses (referred to as drainage Dr in the manuscript) when the rootzone water content exceeds field capacity. This is described in section 2.2.2.

L356-359: Dari et al. (2023) account for evaporative losses in potential ET, similar to AquaCrop, which incorporates ET into the daily calculation of irrigation requirements.

Thank you for noticing this, the sentence has been revised:

L374-376: The retrievals only consist of the irrigation water added to the soil (similar to AquaCrop), therefore not considering runoff or interception losses (included in Noah-MP).

L359-360: While assuming a 100% irrigated fraction for irrigated pixels could lead to overestimation of irrigation, this does not align with the reported values. Furthermore, it may be inconsistent with the approach described in L345–L347, where a conveyance efficiency factor is applied to adjust Noah-MP estimates to match management data.

Compared to Dari et al. (2023), we now acknowledge that this could in theory lead to an overestimation:

L376-377: Also note that our simulations assumed an irrigated fraction of 100% for irrigated pixels, which is likely not the case in reality and should therefore yield an overestimation compared to Dari et al. (2023).

Regarding L345-347, note that the 500-600 mm/year benchmark value has been computed over the irrigated area. This has been clarified in the revised manuscript:

L320-322: Over the entire Po Valley, the average irrigation water use reported by water management agencies is around 500-600 mm yr⁻¹ computed for varying irrigated areas (https://suwanu-europe.eu/wp-content/uploads/2021/05/State-of-play_Po-River-Basin-Italy.pdf, last access 3 February 2026).

L505-506: The conclusion that both models approximate average irrigation rates after accounting for losses seems overstated, given substantial biases. However, the results do not fully support this assertion. Significant errors arise not only from the accounted losses but also from assumptions inherent to each model, including the overestimation of irrigated areas noted above. Importantly, both models apply the same irrigation rule—filling the soil water storage once a critical soil moisture threshold (~50% of TAW) is reached, effectively simulating optimal irrigation, which constitutes a major source of bias.

We agree that the fixed soil moisture threshold leads to unrealistic irrigation behavior at short temporal aggregation scales. At longer time scales, such as annual totals, the large irrigation applications from Noah-MP can result in overestimation compared to AquaCrop. However, we find that differences in the representation of irrigation losses constitute the primary driver of discrepancies in annual irrigation amounts between the models (as shown in Fig. 7).

L520: Add reference for SM_WP and SM_FC being tuned to obtain accurate ET fluxes.

The reference has now been added.

L549: [...] were defined to obtain accurate fluxes (ET), in a mosaic of different land covers (Chen and Dudhia, 2001).

L522: It is unclear how the use of ensembles can better represent irrigation. Moreover, it feels out of context.

We agree with the reviewer that this sentence might be out of context. We removed it.

L570-571: Note that irrigation has been evaluated at finer scales (weekly/bi-weekly) by Dari et al. (2022, 2023), Laluet et al. (2024), and Olivera-Guerra et al. (2023).

Thank you for specifying this. We corrected the sentence:

L604-606: In regions with more intense and larger irrigation networks (e.g. the Ebro basin in Spain), pumping data or larger monitoring systems can serve as a benchmark, ~~but this only allows~~ to validate coarse-scale (~~temporal and spatial~~) irrigation estimates, e.g. ~~monthly estimates~~ at the district scale, as performed in Dari et al. (2023).

L575-576: Highlight Olivera-Guerra et al. (2023)'s approach for calibrating temporal dynamics of irrigation at the district scale, which is compatible with large-scale applications using either crop models or land surface models.

We now highlight this approach:

[L619-621: Nevertheless, promising approaches are emerging that use models and/or remote sensing to infer actual irrigation volumes \(Olivera-Guerra et al., 2023; Dari et al., 2023; Laluet et al., 2024\), demonstrating the potential of these products.](#)

This manuscript presents a timely and well-structured comparison of AquaCrop and Noah-MP, making a valuable contribution by bridging crop and land surface modeling perspectives. The study has strong potential, but the conclusions are currently weakened by the use of heavily aggregated (15-day) ET and SSM comparisons and by limited discussion of uncertainties in satellite-based reference datasets. Expanding the uncertainty analysis, exploring multiple aggregation windows, and offering explicit recommendations to the irrigation community would substantially strengthen the paper.

References:

Dari et al. (2022) : <https://doi.org/10.1016/j.agwat.2022.107537>

Dari et al. (2023) : <https://doi.org/10.5194/essd-15-1555-2023>

Laluet et al. (2024) : <https://doi.org/10.1016/j.agwat.2024.108704>

Olivera-Guerra et al. (2018) : <https://doi.org/10.1016/j.agwat.2018.06.014>

Olivera-Guerra et al. (2023) : <https://doi.org/10.1016/j.agwat.2022.108119>

We thank the reviewer again for the careful and supportive review and trust that all comments have been addressed.

Reviewer 2

General Assessment

This manuscript is well-written, logically structured, and makes an important contribution by comparing AquaCrop and Noah-MP. This work effectively bridges the crop water and land surface modelling communities, providing valuable insights into the representation of irrigation, soil moisture, vegetation, and ET at both basin and field scales. The use of multiple observational and satellite products adds significant robustness. Overall, the paper is of high quality and is generally suitable for publication, requiring minor to moderate revisions focused on strengthening methodological justifications and expanding the interpretation of specific results.

We thank the reviewer for their time and supportive suggestions and give a point-by-point answer below.

Major comments

Given that Section 3.2 focuses on field-based evaluation, yet Section 4.2 later refers to *in situ* irrigation data as “sometime unreliable”, the authors must enhance Section 3.2. A detailed overview of the field data used for comparison is needed, specifically detailing measurement or estimation protocols and the associated uncertainties. This justification supports why these field measurements are considered more reliable than the satellite retrievals, especially considering the better correlation observed with the model simulations. Alternatively, if the data reliability, protocols, and uncertainties are described in a published document or paper, a precise citation to that source should be included with a summary in Section 2.3.4

More details are now given on the *in situ* data described in Section 2.3.4:

*L310-319: Field-based irrigation data is available for three sites and was collected for the ESA IRRIGATION+ project by the Canale Emiliano Romagnolo (CER) consortium, and has been used as benchmark in several studies (Dari et al., 2023; Modanesi et al., 2021a, 2022; Le Page et al., 2023). The Budrio fields consist of five experimental plots within one LIS pixel, covering three irrigation seasons from 2015 through 2017 (top white dot in Figure 1), with the most common crops being tomatoes and maize using drip and sprinkler irrigation. The daily irrigation rates (in mm d⁻¹) were spatially averaged to compare with the irrigation estimates from the models. The Faenza fields (bottom white dot in Figure 1) are separated into two districts: Faenza San Silvestro (2.9 km², covering 3 LIS pixels), and Faenza Formellino (7.6 km², 8 pixels). These *in situ* irrigation rates cover 2016 through 2021 with pear and kiwi as dominant crops using mainly drip irrigation. In this case, the LIS output of multiple pixels was averaged for the fields in each district. Irrigation is evaluated for the months of March through September and is temporally averaged from weekly to seasonal time intervals.*

We also added the possibility of having reliable experimental fields as in situ data:

L601-604: Field-scale irrigation records are typically collected through surveys and agreements with farmers, or alternatively from experimental fields, which are relatively rare. Consequently, field-level in situ irrigation data are limited (Foster et al., 2020) and may also be unreliable, for example when farmers misreport irrigation dates or applied volumes.

Minor comments

Since the acronym TAW is first introduced in the caption of Figure 1, please provide its full definition and briefly explain its significance upon this first appearance.

We now spell out the acronym in the caption of Figure 1:

[...] (c): maps of the total available water (TAW) [mm m^{-1}] (for a rootzone depth of 1 m) derived from the soil hydraulic parameters of each model with Equation 1, and for irrigated cropland only.

For Figures 2 and 3, I recommend employing shared ranges and colorbars (at least per variable) to allow the reader to accurately compare differences between the datasets.

The main message of both figures is to show the difference in spatial patterns, and the mean values. We therefore believe that the difference in color scales is appropriate.

At Line 355, the manuscript states that the Dari's dataset does not include evaporative losses. However, the subsequent methodology utilizes evapotranspiration (ET) rates in the water balance to estimate irrigation. Its implications must be explicitly considered when comparing the results against AquaCrop and Noah-MP.

Thank you for noticing this. This sentence has been revised to:

L374-376: The retrievals only consist of the irrigation water added to the soil (similar to AquaCrop), therefore not considering runoff or interception losses (included in Noah-MP).

At Line 381, I suggest including a brief explanation for the decision to evaluate the additional DMP data specifically during the months from January to June.

This was already specified in the methodology and now further clarified:

L294-296: Similarly to the SSM, the vegetation was evaluated for the months March through September, and also separately for the first half of the year (January-June) corresponding to the period when the AquaCrop CC increases (before reaching a plateau in the summer) as in de Roos et al. (2024).

It is also partly repeated in the Results for clarity:

L468-499: [...] corresponding to the period when the AquaCrop CC increases (before reaching a plateau in summer).

I think the paper could close with an overview of implications on this paper on the future investigations, regarding the utility of estimating the irrigation and the gap between crop and land surface models.

Thank you for your suggestion, we now explicitly refine how these modeling communities can help each other. In Section 4 “Overview, limitations, and future perspectives”.

L561-564: To better represent irrigation (and crop processes), the LSM community is transitioning toward the use of dedicated submodels, such as Noah-MP-Crop (Liu et al., 2016a), in which crop-specific processes and parameters are explicitly represented. This shift inherently moves LSMs toward finer spatial scales (or tiling), helping to bridge more detailed crop model processes with the broader LSM framework.

L575-578: Compared to LSMs, the role of crop models is distinct: they do not aim to simulate the full suite of land surface processes. Yet advances in this research field are increasingly being incorporated into coarser-scale LSMs, provided they are generalizable enough, which is something crop models can further refine by learning from the LSM community.

In the “Conclusions”, we also highlight the contributions from both modeling communities:

L548-550: Although their original roles are distinct, both communities can learn from each other: LSMs increasingly take up processes from crop models, while the crop modeling community can in turn draw insights from the generalization strategies used in LSMs.

Reviewer 3

This manuscript presents a comparison between AquaCrop and Noah-MP within NASA's LIS framework for irrigation estimation over the Po Valley. The integration of AquaCrop into LIS is very interesting and innovative contribution, opening new opportunities for the community to apply AquaCrop in regional, high-resolution hydrological contexts.

We thank the reviewer for their time and constructive feedback. We provide a point-by-point response to all comments below.

In my opinion (sorry), at the same time, several conceptual and methodological aspects limit the scientific strength of the study in its current form. In my opinion the most critical point is the crop representation strategy, which strongly influences irrigation and land-surface processes. The use of a generic transplanted crop as the representative vegetation for the Po Valley is not adequately justified. Since AquaCrop is inherently crop-specific and phenology-driven, this choice strongly influences canopy development, rooting depth evolution, timing and magnitude of irrigation. In my opinion, considering that only summer crops (and sometime Alfalfa) are irrigated in Po Valley, my proposal is to use a realistic summer crop (e.g., maize), or justify in detail why this was not feasible.

Re-running simulations with a realistic crop in my opinion can greatly enhance the quality and the interest of the comparative analysis.

We thank the reviewer for this important comment and fully agree that crop representation strongly influences irrigation dynamics and land-surface processes. The manuscript is intentionally interdisciplinary and aims to compare two fundamentally different modeling frameworks: a crop-specific, phenology-driven model (AquaCrop) and a land surface model with a generic vegetation representation (Noah-MP, which cannot be run with a specific crop). To avoid ambiguity, we further clarified the scope and objectives of the study in the Introduction, explicitly positioning the work as a model intercomparison focused on conceptual and structural differences rather than crop-specific optimization over the Po Valley.

L81-85: Both LSMs and crop models have different original purposes and scales, but they can serve the same application, namely, to estimate irrigation regionally. The objective of this study is to compare how differences in model structure and process representation between a crop model and an LSM translate into differences in irrigation estimates and related variables at regional (basin) and pixel scales (field-based evaluation). The Po Valley (Italy) is used as an illustrative study domain to examine the model behavior and process differences.

The use of a generic transplanted crop in AquaCrop is a deliberate choice to maintain conceptual consistency with the land surface model rather than to reproduce a specific cropping system in the Po Valley. Noah-MP is not designed to represent explicit crops

such as maize or crop-specific phenology and management. AquaCrop has previously been applied at regional scale using generic crop representations (de Roos et al., 2021, 2023, 2024; Busschaert et al., 2022). We therefore left model-intrinsic structures and parameterizations unchanged, as harmonizing parameters to minimize differences would undermine the scientific purpose of the comparison. While this choice affects canopy development, rooting depth, and irrigation timing, these differences are central to the research question addressed in this study.

Intermediate-minor concern

Parametrization: The two models rely on different soil hydraulic parameterizations. Many differences in simulated irrigation or soil moisture may therefore reflect parameter space differences rather than model process differences. A clearer discussion would help guide interpretation.

Indeed, the models rely on different soil hydraulic parameters but the soil texture classes are identical for both models (extracted from the Harmonized Soil World Database). This is explained in the manuscript and can also be seen in Fig. 1b and c which present the same spatial patterns. These lines discuss the difference in soil hydraulic parameters:

L143-155: The soil texture classes were assumed to be homogeneous for the whole profile, identical for both models, and derived from the Harmonized Soil World Database (HWSD) v1.21. However, each model uses its own lookup tables to map the mineral soil texture to different estimates of associated soil hydraulic parameters (SHP), i.e. the water content at saturation θ_{sat} , at field capacity θ_{FC} , at wilting point θ_{WP} , and the saturated hydraulic conductivity, K_{sat} . Unlike other studies that improved the soil hydraulic parametrization for bucket-type models (e.g. Romano et al., 2011, 2025), in this study, the prescribed SHPs are used for each soil texture class without considering spatial variation, as is commonly done for large-scale simulations (De Lannoy et al., 2014; Kishné et al., 2017). Furthermore, soil organic carbon is not explicitly accounted for. For AquaCrop, indicative values of the SHPs are provided for each soil texture class (Raes et al., 2023), which are inherited from field-based applications (higher θ_{WP} and lower θ_{FC}), whereas Noah-MP has been developed and calibrated with different θ_{FC} and θ_{WP} values, defined by Chen and Dudhia (2001). More specifically, the Noah-MP SHPs are based on Cosby et al. (1984) and further adapted to intentionally increase total available water (TAW; also referred to as plant-available water) in the context of large-scale simulations by Chen and Dudhia (2001). Because of the inherent characteristics of the models, each model requires its own parameter set, which is further discussed in Section 4.2.

The 45% irrigation threshold, coming from earlier studies, may not be optimal for AquaCrop or for the selected generic crop. Some justification, sensitivity analysis, or recalibration would strengthen the results (note that in several models the threshold is below 33% of the TAW)

This is now clarified in the text:

L209-213: In line with the Noah-MP configuration, the irrigation threshold is set to 45% of the TAW. In AquaCrop, the irrigation threshold is defined in terms of depletion of the rootzone readily available water (RAW), which is itself a fraction of TAW determined by the crop-specific p-factor. For the generic C3 crop used in this study, the p-factor is set to the default value 0.5, resulting in a depletion threshold of 110% of RAW and allowing for limited water stress prior to irrigation. This configuration ensures conceptual consistency between the two modeling frameworks rather than optimization for a specific crop.

The use of GVF and CC thresholds produces growing seasons that do not always correspond to real agricultural calendars in the Po Valley, mainly because there are (I imagine frequently) cases with cells that have inside winter and summer crop. Some problem still can arise when a winter crop like Barley or Triticale (not irrigated) are harvested early for forage and after is followed by maize

We define the growing season as the period when irrigation is allowed (irrigation is not relevant for winter crops in the region). This is now clarified throughout the manuscript.

L134-135: The model-specific settings and the definition of the growing season, defined as the period when irrigation is allowed, are presented in the next subsections.

Table 1: [...] Note that the growing season is defined as the period when irrigation is allowed.

L249: For both model setups, a dynamic start of the growing season (time window when irrigation is allowed) was defined.

L734-735: The growing season corresponds to the period when irrigation is allowed (can be triggered in the model).

The manuscript is very long and at times overly detailed. Several paragraphs could be substantially condensed without loss of information. Some sections (especially the Results) are verbose and occasionally difficult to follow.

To shorten the length of the manuscript, we have moved previous Section 3.1.1 “Growing season” to the new Appendix C. The current structure and depth results directly from integrating extensive revisions requested in the first review round, and we believe it is an appropriate and necessary level of detail for an interdisciplinary audience to avoid misunderstandings.

In my opinion several graphs (the maps) are difficult to interpret due to different scales for the same variables. Some caption is too synthetic and full of acronyms, some other is too long (I think that we don't need the explication of what is a boxplot in the caption, and what is the reason to omit the outliers ?

Figures shown with different color scales are intended to emphasize spatial patterns and mean values rather than enable direct quantitative comparison across panels. We therefore deliberately retain different scales to better highlight spatial variability within each variable.

Regarding figure captions, we have revised them for improved clarity and consistency. Acronyms are spelled out in Figure 1, and the remaining acronyms (e.g., CGLS DMP, SMAP SSM) are used consistently throughout the manuscript and are defined in the text. We therefore consider them sufficiently clear.

We agree that captions should remain concise. However, we consider it important to explicitly define the boxplot statistics, as boxplots are not always drawn following a single standard (e.g., differences in whisker definitions or interquartile ranges). To address this concern, we now provide the boxplot definition only once (in Figure 3, previous Fig. 4) and have removed the detailed explanation from Figure 4 (previous Fig. 5).

Finally, outliers are not shown in the boxplots to improve readability given the large number of data points; this choice does not affect the underlying statistics shown.

I also attach the annotated PDF, where I have included several rough (apologies) but hopefully useful comments and margin detailed notes that may help the authors to understand my criticism and to improve the manuscript.

We thank the reviewer for the time and effort invested in this review. We have responded to all comments above and trust that these clarifications improve the understanding of the manuscript.

References:

de Roos et al. (2021): [10.5194/gmd-14-7309-2021](https://doi.org/10.5194/gmd-14-7309-2021)

de Roos et al. (2023): [10.1016/j.rse.2023.113621](https://doi.org/10.1016/j.rse.2023.113621)

de Roos et al. (2024): [10.1029/2024JG008231](https://doi.org/10.1029/2024JG008231)

Busschaert et al. (2022): [10.5194/hess-26-3731-2022](https://doi.org/10.5194/hess-26-3731-2022)

Reviewer 4

Review on paper « 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 » by Busschaert et al. submitted for publication in HESS

This study aims to compare the ability of a land surface model and a crop model to predict irrigation at the regional scale and other variables related to the water cycle including soil moisture, evapotranspiration and dry matter production. The paper is well written and well organized and is of interest for the HESS community. Nevertheless, I have some major concerns listed below.

[We thank the reviewer for their time and thorough review. We provide a point-by-point answer to all comments below.](#)

Positioning of the research question

The objective of the study is described as « estimating irrigation at the regional scale » but it is not clear what's the purpose behind estimating irrigation at the regional scale. If it is for water management, the model should relate the water withdrawal for irrigation to the surface and ground reservoir. If it is for improving the land surface condition for atmospheric models, spatial resolution of 1 km² is probably too fine. It cannot be for irrigation scheduling as the resolution is too coarse ... Please elaborate on this point. Still considering the positioning of the research question, what's the objective of comparing a Land Surface Model to an agronomic model as many LSM, thanks to recent developments, include irrigation modules quite similar in their basic principle to aquacrop (refill the reservoir when available water fall below a given threshold) ? By contrast, I see significant difference in the way the crop growing processes are represented and I was wondering if the contrasting results were not related more to the difference in terms of dynamic vegetation than in terms of irrigation prediction.

[We thank the reviewer for this constructive concern. We clarified the objective of the study.](#)

[L81-85: Both LSMs and crop models have different original purposes and scales, but they can serve the same application, namely, to estimate irrigation regionally. The objective of this study is to compare how differences in model structure and process representation between a crop model and an LSM translate into differences in irrigation estimates and related variables at regional \(basin\) and pixel scales \(field-based evaluation\). The Po Valley \(Italy\) is used as an illustrative study domain to examine the model behavior and process differences.](#)

[We added these lines to the Introduction to better position the study and clarify its objective.](#)

L33-35: Estimating irrigation from models and remote sensing has become a shared objective across hydrology (for water demand assessments; Döll and Siebert, 2002), agricultural water management (for improved decision making and planning; Foster et al., 2020), and land-atmosphere research (because of the strong feedbacks that irrigation induces in the climate system; Yao et al., 2025).

We also note that modeled net irrigation can be converted to gross withdrawals if needed for specific applications:

L79-80: The net requirements can be scaled with efficiency factors to give estimates of water withdrawals (Döll and Siebert, 2002).

Methodological aspects of the comparison

The authors choosed to use the models « out of the box » with their own parameterization and input variables without trying to adjust the inputs and parameterization in order that both models runs in close configuration seems questionable to me. Is it not an obvious result that the model provide with quite different irrigation estimates considering the large differences in crop phenology, soil moisture dynamic and thus probably root depth dynamic,? It is even not necessary to run the models to show that irrigation estimates with NOAH will be much higher than with Aquacrop. Having a look to the TAW values for both model is sufficient (as written by the authors at L470).

We understand the reviewer's concern and agree that differences in crop phenology, soil moisture dynamics, and rooting depth inevitably lead to different irrigation estimates between Noah-MP and AquaCrop. However, this outcome is central to the objective of the study rather than a limitation.

The purpose of this work is not to harmonize the two models, but to compare how two fundamentally different modeling frameworks represent irrigation when applied at regional scale. We harmonized what should be harmonized for a fair intercomparison (forcing, soil map, irrigation mask, ...) while leaving model-intrinsic structures and their parameters untouched (generic crop representation, soil hydraulic lookups, ...).

We agree that, given the Noah-MP soil hydraulic parameterization (larger TAW), larger irrigation amounts are expected at short temporal aggregation intervals. However, irrigation differences cannot be inferred from TAW alone. At seasonal time scales, water losses play a dominant role, as demonstrated by the water balance analysis in Fig. 7, which shows that differences in simulated losses strongly influence net irrigation totals.

The value of this comparison lies not in the magnitude of irrigation estimates alone, but in identifying how model structure and process representation drive these differences.

Considering the comparison with the Dari data set, I see a major concern as well. One of the main conclusion of the study is that the fact that both model irrigation estimates tend to correlate better with each other than with the satellite data sets by aggregating at larger

time scale (the best correlation is obtained for correlation computed over the whole summer) demonstrate that the model estimates are more consistent than the satellite data set of Dari. To my opinion, it just show that the atmospheric evaporative demand, governed by the forcing meteorological data set that is the same for both model only ... I agree that satellite irrigation estimates are really uncertain but uncertainties are high for model estimates as well (forcing, irrigation map, soil characteristics ...) and this study does not demonstrate that the model estimates are of better quality for the reasons above. The next section showing the comparison to in situ irrigation is not sufficiently convincing by lack of important information. Could you tell the reader a little bit more about the representativity of the in situ irrigation data, we don't know how the irrigation data were gathered and we don't even know the crops that were grown on the sample fields. Could you provide these information ?

The reviewer is correct that irrigation requirements simulated by both models are largely governed by atmospheric evaporative demand and therefore by the shared meteorological forcing (MERRA2). This is precisely the point we aim to demonstrate. The increased agreement between model-based irrigation estimates at larger temporal aggregation scales reflects their common forcing, rather than indicating higher accuracy of the model estimates relative to the Dari et al. (2023) dataset.

We do not claim that the model estimates are superior to satellite-based irrigation products. Instead, the comparison highlights a fundamental conceptual difference between modeled irrigation, which is driven by soil moisture deficit thresholds, and observation-based datasets, which, at short time scales, are more likely to reflect real-world irrigation practices based on scheduling rather than instantaneous moisture deficits, as stated in the manuscript

L380-382: Observation-based datasets, such as the one presented by Dari et al. (2023) may better capture the reality at short time scales, where farmers likely irrigate according to a schedule and not depending on moisture deficit thresholds (Pokhrel et al., 2016), and may also face water-use restrictions during drought years.

Regarding the in situ irrigation data, we have expanded the manuscript to provide additional information on data collection, field representativity, and crop types, in line with the reviewer's comment and R2's suggestions, to better support the field-scale evaluation.

L310-319: Field-based irrigation data is available for three sites and was collected for the ESA IRRIGATION+ project by the Canale Emiliano Romagnolo (CER) consortium, and has been used as benchmark in several studies (Dari et al., 2023; Modanesi et al., 2021a, 2022; Le Page et al., 2023). The Budrio fields consist of five experimental plots within one LIS pixel, covering three irrigation seasons from 2015 through 2017 (top white dot in Figure 1), with the most common crops being tomatoes and maize using drip and sprinkler

irrigation. The daily irrigation rates (in mm d⁻¹) were spatially averaged to compare with the irrigation estimates from the models. The Faenza fields (bottom white dot in Figure 1) are separated into two districts: Faenza San Silvestro (2.9 km², covering 3 LIS pixels), and Faenza Formellino (7.6 km², 8 pixels). These in situ irrigation rates cover 2016 through 2021 with pear and kiwi as dominant crops using mainly drip irrigation. In this case, the LIS output of multiple pixels was averaged for the fields in each district. Irrigation is evaluated for the months of March through September and is temporally averaged from weekly to seasonal time intervals.

In addition, the manuscript was at times hard to follow, as (1) important details on model parameterization and validation data are not provided, and simply referring to previous papers is not sufficient. Readers cannot reasonably be expected to spend several days reviewing the literature in order to understand a single study and (2) some choice of model setup appear questionable and are not justified. Finally, there is no clear conclusion apart that more in situ data should be gathered and that the remote sensing community should improve the satellite retrieval of irrigation. Could the authors suggest a few research directions to improve the models?

Providing all model details and parameterization processes would make the manuscript overly long and difficult to follow. As already noted by R3, we therefore aim to strike a balance between the level of technical detail and the overall readability of the manuscript.

We agree that the main conclusion for both modeling communities was unclear and clarify it now in the section “Overview, limitations, and future perspectives”.

L561-564: To better represent irrigation (and crop processes), the LSM community is transitioning toward the use of dedicated submodels, such as Noah-MP-Crop (Liu et al., 2016a), in which crop-specific processes and parameters are explicitly represented. This shift inherently moves LSMs toward finer spatial scales (or tiling), helping to bridge more detailed crop model processes with the broader LSM framework.

L575-578: Compared to LSMs, the role of crop models is distinct: they do not aim to simulate the full suite of land surface processes. Yet advances in this research field are increasingly being incorporated into coarser-scale LSMs, provided they are generalizable enough, which is something crop models can further refine by learning from the LSM community.

In the “Conclusions”, we also highlight the contributions from both modeling communities:

L548-550: Although their original roles are distinct, both communities can learn from each other: LSMs increasingly take up processes from crop models, while the crop

[modeling community can in turn draw insights from the generalization strategies used in LSMs.](#)

To my opinion, the paper has potential for publication but I recommend major revision including significant rewriting.

Detailed comments

Both models were run as sprinkler irrigation (meaning that irrigation is added to rainfall, right ?) whereas surface irrigation is the dominant practice in the study area (L96). This inconsistency raises questions about how representative the simulations are. In addition, part of the discussion focuses on canopy interception differences, which are simulated in Noah but not in AquaCrop; however, given that the prevailing irrigation method is surface rather than sprinkler, it is not clear what the added value of this discussion is. This is one the point among others that seems to complicate the message unnecessarily.

We assume sprinkler-type application for both models because surface irrigation is not implemented in Noah-MP within NASA's Land Information System. In our study, we do not claim absolute realism for Po Valley irrigation practices but to highlight differences in how two modeling frameworks estimate regional irrigation demand under a common and simplified irrigation assumption. We clarified the differences in losses below.

L357-361: AquaCrop aims to estimate a net irrigation amount and only accounts for soil evaporation losses. Noah-MP additionally simulates canopy interception and surface runoff losses, which can occur for sprinkler irrigation. However, neither of the models explicitly represents percolation losses due to irrigation, which can be substantial for surface irrigation systems (dominant irrigation method over the area).

L120 : Why a spin up is needed for Aquacrop as the reservoir is refilled every winter ?

There is no guarantee that the reservoir gets refilled every winter over the entire study domain. To be safe, a spin up is performed. Nevertheless, we agree with the reviewer that this might be unnecessary for the area, especially under irrigated conditions.

It would be interesting to further the comparison with the data set of Dari et al. including the spatial patterns (fig. 2) and on the validation sites (Fig 9, G1, G2)

We agree that extending the comparison with the Dari et al. (2023) dataset would be interesting. However, this comparison is intended as a complementary element rather than a central focus of the study. Given the current length and level of detail of the manuscript, and concerns raised about its complexity (R3), we have chosen not to extend the analysis further.

Legend of figure 1 : explicit SHPs

SHPs have been spelled out in the caption.

Table 1 : explicit CC and GVF

This has been made explicit.

L128 : please justify the bilinear interpolation by providing the elevation values of the study region, at least the range of elevation.

This is now clarified in the manuscript:

L137-141: The forcing data were horizontally interpolated to the model grid using bilinear interpolation. Approximately 15% of the study domain shows elevation differences larger than 200 m relative to the forcing grid, for which vertical downscaling (lapse-rate correction) may be beneficial. However, this correction was not compatible with AquaCrop v7.0 in NASA's LIS and was therefore applied only for Noah-MP.

L140 : « More specifically, the Noah-MP SHPs are based on Cosby et al. (1984) and further adapted to intentionally increase total available water (TAW; also referred to as plant-available water) in the context of largescale simulations by Chen and Dudhia (2001). » Why is it necessary to intentionally increase TAW in the context of large scale simulations ?

The intentional increase in total available water (TAW) in Noah-MP follows the rationale of Chen and Dudhia (2001) to account for subgrid-scale soil moisture heterogeneity in large-scale simulations. At the grid-cell scale of land surface and atmospheric models, soil moisture is spatially heterogeneous, meaning that even when the grid-average soil moisture approaches the wilting point, portions of the grid can still sustain evaporation (Chen et al., 1996). Increasing the effective TAW allows the model to represent this unresolved heterogeneity in a simplified manner, maintaining nonzero evaporation under dry conditions without explicitly resolving subgrid variability. This approach provides a computationally efficient way to represent area-averaged land-atmosphere fluxes at coarse spatial resolution.

L159 : the rooting depth not only depends on the geographic location but also on the type of crops. The sentence should be reformulated.

The sentence has been rephrased to the following:

L172-174: In the region, rooting depth varies with both soil conditions and crop type and typically ranges from 50 to 150 cm, with deeper root zones occurring near the edges of the Po River (Rivieccio et al., 2020).

L170 : the choice of C3 is questionable in this region that is the main producer of maize. I understand that it won't change the conclusion of your study but It doesn't enhance its credibility from an agronomic point of view either. If you prefer considering C3, wheat is sown in autumn and harvested by the end of June. I understand that for this exercise we're

trying to simplify as much as possible by applying the same crop across the whole study area, but why not, at the very least, choose realistic cropping periods?

We acknowledge the reviewer's agronomic concern and agree that maize is the dominant crop in the Po Valley. However, the use of a generic C3 crop is a deliberate choice aligned with the objectives of this study. The aim is not to represent a specific crop or cropping calendar, but to apply a simplified and spatially uniform vegetation representation that is conceptually comparable to Noah-MP, which does not include an explicit crop module allowing crop-specific phenology or planting dates.

The generic C3 crop uses average parameter values intended to represent typical summer crops in terms of canopy development, rooting depth, and transpiration capacity, and has been successfully applied in previous large-scale AquaCrop studies over Europe (de Roos et al., 2021, 2023, 2024; Busschaert et al., 2022). Adopting a maize-specific parameterization or realistic cropping periods would introduce crop-specific assumptions that cannot be mirrored in Noah-MP and would therefore undermine the balance of the intercomparison.

L201 : January 1st to the end of March (80 calendar days) is more than 80 days.

This has been corrected in the manuscript:

L217-218: The rootzone depth RZ [m] is defined by the crop parameters and starts at 0.1 m on January 1st to reach its maximum value (RZ_{max}) of 1 m after 80 calendar days.

L203 : « except for the radiation transfer scheme » please explain

We added the following sentence to the manuscript:

L225-226: The radiation transfer had to be changed accordingly to make it compatible with the dynamic vegetation option.

L204 : « The latter enables a dynamic definition of the growing season in line with AquaCrop » At least explain how LAI (which is the prognostic variable, right ?) is computed and the basic principle for the start of the growing season. Is there a seeding date defined ? (January 1st like Aquacrop ?). We finally learn at L402 that harvest is not modeled but what about seeding ? Is there a photosynthesis module and growth start when carbon assimilation is enough to maintain existing tissues and fill some carbon reservoirs ?

In contrast to AquaCrop, there is no sowing/planting date defined in Noah-MP and the vegetation growth is managed through the carbon assimilation. This is now clarified in the text:

L223-225: In contrast to AquaCrop, Noah-MP does not prescribe a sowing/planting date; vegetation growth and LAI evolve dynamically from carbon assimilation, with the growing season starting when environmental conditions allow net carbon gain (Niu et al., 2011).

L218 : the discretization of the soil vertical profile is much finer for Aquacrop than for NOAH. How Irrigation applied is computed in Noah ? For instance, if root depth is larger than the third layer, let's say 65 cm, do you refill the four layer including the last one going down to one meter entirely ?

For both models, the irrigation application dose is computed based on the dynamically evolving root zone rather than fixed soil layers. In Noah-MP, irrigation water is applied to fill the soil moisture deficit within the dynamic root zone, meaning that only the fraction of a soil layer intersecting the root zone is refilled when the root depth does not coincide with layer boundaries. We now explicitly describe this irrigation application logic for Noah-MP by referring to the Irr_{appl} equation:

L246-247: The Irr_{appl} is computed as in Equation 5 and is equally distributed at each model time step (15 min) and applied from 06:00 to 10:00 AM (local time) and added to the precipitation.

L230 : « This threshold is defined as 40% » This is the threshold for the start of the growing season or the threshold when irrigation can be triggered ?

This has been clarified now:

L249-252: In Noah-MP, the growing season, or the time when irrigation can be triggered in the model, was initially described by Ozdogan et al. (2010) as the period where the GVF is larger than 40% of the range between the minimum and maximum GVF (GVFmin and GVFmax). This definition has been used in several studies (Lawston et al., 2017; Lawston-Parker et al., 2023; Modanesi et al., 2021b; Sharma et al., 2022).

2.3.3 Evapotranspiration product / I understand that you were looking for a publicly available data set and you didn't wanted to recompute an evapotranspiration product on your own. Nevertheless, SenET is based on 1km² surface temperature product derived from Sentinel-3 and downscaled with Sentinel-2 at 100 m and you average again to compare to your 1km grid point Please, at least, point out this aberration.

This is now made transparent in the manuscript:

L299-302: The ET is derived using the two-sourced Energy Balance model that downscales the original 1-km Sentinel-3 land surface temperature using Sentinel-2 surface reflectance. For consistency with the model grid, the data were reprojected by spatial averaging. As a result, the ET estimates remain primarily constrained by the original 1-km Sentinel- 3 land surface temperature; however, the re-aggregation may introduce some uncertainty.

L240 : I don't understand the first half of this sentence « . The GV F in Noah-MP is based on the dynamically simulated LAI via an exponential relation (Equation 8) whereas the CC in AquaCrop is the driving variable and is based on the climatic and environmental conditions » in Noah-MP, GVF is directly derived from the dynamical variable LAI that

depends as well on the climatic and environmental conditions, right ? The message of this sentence is not clear to me.

We clarified this sentence:

L259-261: In Noah-MP, GVF is diagnostically derived from the prognostic LAI through an exponential relationship (Equation 8), whereas in AquaCrop, the CC is the primary state variable driving surface fluxes and is explicitly controlled by crop growth parameters and stress responses.

L245 : « Note that all these reference products have their uncertainties, in particular over irrigated areas » Is this sentence really useful considering that all data are uncertain all the more when considering remote sensing products. Could you explain why the data uncertainties should be higher on irrigated areas ?

We believe it is useful to state that remote sensing products do typically have large uncertainties over irrigated areas. This is described in detail in Section 4.4.

L251 : Specify that these are the model estimates.

This is now specified:

L270: The surface soil moisture (SSM) model estimates were evaluated [...]

Fig3. Please use the same color scale

Figures shown with different color scales are intended to emphasize spatial patterns and mean values rather than enable direct quantitative comparison across panels. We therefore deliberately retain different scales to better highlight spatial variability within each variable.

L287 : please explain how these irrigation data were gathered and provide with the crops and irrigation technic of each fields.

This paragraph has been revised

L310-319: Field-based irrigation data is available for three sites and was collected for the ESA IRRIGATION+ project by the Canale Emiliano Romagnolo (CER) consortium, and has been used as benchmark in several studies (Dari et al., 2023; Modanesi et al., 2021a, 2022; Le Page et al., 2023). The Budrio fields consist of five experimental plots within one LIS pixel, covering three irrigation seasons from 2015 through 2017 (top white dot in Figure 1), with the most common crops being tomatoes and maize using drip and sprinkler irrigation. The daily irrigation rates (in mm d⁻¹) were spatially averaged to compare with the irrigation estimates from the models. The Faenza fields (bottom white dot in Figure 1) are separated into two districts: Faenza San Silvestro (2.9 km², covering 3 LIS pixels), and Faenza Formellino (7.6 km², 8 pixels). These in situ irrigation rates cover 2016 through 2021 with pear and kiwi as dominant crops using mainly drip irrigation. In this case, the

LIS output of multiple pixels was averaged for the fields in each district. Irrigation is evaluated for the months of March through September and is temporally averaged from weekly to seasonal time intervals.

L296 : Please provide temporal and spatial resolution of the irrigation data set as well as the time period of availability.

This information has been added to the manuscript:

L322-323: [...] and are available weekly at a 1-km spatial resolution from January 2016 through July 2020.

L297 : The first part of the sentence « As the first regional irrigation product for the Po Valley, it is subject to notable uncertainties, especially since » is useless without providing an assessment of this data set.

We think that this sentence is valid because we point out that the dataset was only validated using the Faenza fields. We rephrased the sentence and hope it clarifies our point.

L324-325: As the first regional irrigation product for the Po Valley, it is subject to notable uncertainties, especially since the algorithm was only validated over the Faenza fields.

L335 : please show the spatial patterns of precipitations and radiation. It would be interesting for comparison purpose and they can be easily inserted within an appendix.

Maps of precipitation and ET (radiation driven) are shown in De Lannoy et al. (2024).

L349-351: The main drivers for both models are radiation and precipitation, presenting similar spatial patterns (not shown here but detailed in De Lannoy et al., 2024).

L341 : « Noah-MP includes the simulation of interception » is this really problematic as most of irrigated field are through surface irrigation ?

We rephrased the sentence to

L357-361: AquaCrop aims to estimate a net irrigation amount and only accounts for soil evaporation losses. Noah-MP additionally simulates canopy interception and surface runoff losses, which can occur for sprinkler irrigation. However, neither of the models explicitly represents percolation losses due to irrigation, which can be substantial for surface irrigation systems (dominant irrigation method over the area).

L341 : «Noah-MP includes the simulation of interception and runoff losses that AquaCrop does not simulate, and which are important for sprinkler irrigation ». Why runoff is important for sprinkler irrigation ?

Runoff has been shown to cause substantial losses in sprinkler irrigation systems (Luz and Heermann, 2005).

L351 : quite obvious as both models have the same meteorological forcing, right ?

The sentence has been rephrased:

L368-369: First, as expected given the shared meteorological forcing, the temporal evolution of median irrigation is similar in both models, [...]

L363-364 : knowing if there have been some administrative order restricting agricultural water use in 2021 could strengthen the credibility of the Dari dataset, which shows a decrease in irrigation water demand in that particular year.

We did not find evidence that there were water restrictions for irrigation in 2021 over the Po valley.

L364 : I would add « and could face water use restriction during drought year »

This has been added.

L381-382: [...]and not depending on moisture deficit thresholds (Pokhrel et al., 2016), and may also face water-use restrictions during drought years.

L389 : « The irrigation practices applied in rice fields differ strongly from the sprinkler irrigation modeled by AquaCrop and Noah-MP » But I thought that the main irrigation technics was surface irrigation ?

This has been rephrased:

L406-408: The irrigation practices applied in rice fields differ strongly from the sprinkler-type irrigation assumed in AquaCrop and Noah-MP, as basin irrigation with prolonged ponding deviates even more strongly from these assumptions than other surface irrigation methods.

L403 «The choice of vegetation module options in this study may have resulted » again, this is not a report for the NOAA team but a scientific article for the scientific community. What are the different options ?

We believe that detailing the different Noah-MP vegetation options is beyond the scope of this study. The sentence is intended only to acknowledge that model results can depend on vegetation module choices, and we therefore refer interested readers to Li et al. (2022) for further details.

L431 «Note also that in Noah-MP, the irrigation water lost through runoff is not compensated, meaning that field capacity is never reached in the model after an irrigation event» It is not clear to me what runoff are you referring. Is it runoff due to infiltration excess ? Meaning that the irrigation rate is too strong with regards to the soil infiltration capacity ? Is it realistic ?

The reviewer is right to point out that this is not realistic. The text was rewritten as follows:

L456-460: In Noah-MP, surface runoff during irrigation is generated through a runoff scheme (Yang and Dickinson, 1996), which is appropriate for representing infiltration-excess runoff from rainfall at the grid scale. However, when applied to large irrigation applications, as shown in Section 3.2, the resulting runoff losses are not compensated by additional irrigation input, preventing soil moisture from reaching field capacity and leading to an unrealistic representation of field-scale irrigation.

Fig 9 : could you please show the satellite irrigation retrieval as well.

The comparison with the Dari et al. (2023) dataset is intended as a complementary element rather than a core objective of the study. Given the level of detail already included in the manuscript and to honor the request of other reviewers to keep it shorter, we prefer not to extend the evaluation further.

L500 : « Overestimations were also found in other studies » of what by which model ?

The studies are specified in the manuscript: Modanesi et al. (2022) and Dari et al. (2023).

L501 : « due to the dominance of fruit trees » meaning that the in situ data are gathered on fruit trees ?

The section on the in situ data description has been revised to clarify this.

L500-502 : « Overestimations were also found in other studies (Dari et al., 2023; Modanesi et al., 2022) and are likely due to [...] localized methods (drip irrigation), which are more efficient than sprinkler irrigation / Probably but it can be simulated by aquacrop through the fw (wet fraction) factor.

The reviewer is right that this could be simulated by AquaCrop, but for the sake of consistency with Noah-MP, this is not explored in this study.

L505 : « and application losses that are included in water records » I don't understand the importance of mentioning the water losses through transportation as you compare your model predictions to in situ irrigation data and satellite retrieval that both represent the water really applied at the field scale.

This sentence was referring to the comparison with reported water records. This has now been rephrased for clarity.

L535-537: By design, AquaCrop simulates net irrigation amounts and Noah-MP estimates irrigation amounts with minimal application losses, and both do not account for larger transportation and application losses that are included in the reported water records (on average 500-600 mm yr⁻¹).

L570 : « but this only allows to validate coarse-scale (temporal and spatial) irrigation estimates » In relation to my general comment about the positioning of the research

question, this could be an acceptable objective but it needs to be explained in the introduction.

The objective has been clarified, as stated in the answer to the reviewer's general comments.

Regarding this specific comment, we decided to slightly rephrase the sentence to avoid any confusion.

L604-606: In regions with more intense and larger irrigation networks (e.g. the Ebro basin in Spain), pumping data or larger monitoring systems can serve as a benchmark, but this only allows to validate coarse-scale (temporal and spatial) irrigation estimates, e.g. monthly estimates at the district scale, as performed in Dari et al. (2023).

L576: « First, the native spatial resolution of the remote sensing products (relatively coarse compared to fields) complicates the detection of irrigation signals (in the SSM or vegetation; Ozdogan et al., 2010) » I have two comment on this sentence. The size of your grid point is 1km² similar to the resolution of the Dari dataset so probably much higher than the typical field scale as well. Once again, what is the objective of this study ?

The objective has been clarified, as stated in the answer to the reviewer's general comments.

We agree that 1 km² is typically much larger than individual fields and comparable to the resolution of the Dari et al. (2023) dataset. Our intent is to emphasize that, at this spatial scale, land cover mixing within a pixel remains substantial, which complicates the detection of irrigation signals in soil moisture or vegetation observations. We have therefore clarified the sentence as follows:

L609-611: First, the native spatial resolution of the remote sensing products (relatively coarse compared to individual fields) leads to substantial heterogeneity within a pixel, both due to mixed land cover and variability within croplands, which complicates the detection of irrigation signals in soil moisture or vegetation observations (Ozdogan et al., 2010).

We would like to thank again the reviewer. We answered all comments and edited the manuscript accordingly.

References:

Chen et al. (1996): <https://doi.org/10.1029/95JD02165>

de Roos et al. (2021): [10.5194/gmd-14-7309-2021](https://doi.org/10.5194/gmd-14-7309-2021)

de Roos et al. (2023): [10.1016/j.rse.2023.113621](https://doi.org/10.1016/j.rse.2023.113621)

de Roos et al. (2024): [10.1029/2024JG008231](https://doi.org/10.1029/2024JG008231)

Busschaert et al. (2022): [10.5194/hess-26-3731-2022](https://doi.org/10.5194/hess-26-3731-2022)

Luz and Heermann (2005): <https://doi.org/10.1016/j.agwat.2004.09.013>