We thank both Referees for their review and comments. 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.

Both Referees #3 and #4 raised concerns about the parameterizations affecting the comparison of the two models, particularly regarding vegetation representation (Referee #3) and soil hydraulic properties (Referee #4). We therefore now clarify that this manuscript presents a process-oriented intercomparison in which model-intrinsic structures and parameterizations are intentionally preserved, rather than a strictly controlled comparison with harmonized settings. We hope this helps with the clarity of the manuscript.

In the introduction L82-86: *The objective of this study is to perform a process-oriented intercomparison of two models to assess how differences in model structure, process representation, and model-intrinsic parameterizations 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).*

Referee #3

Thank you for your careful revision of the manuscript and for the detailed responses to the reviewers' comments. The rationale for comparing two very different model types is now quite well justified in the revised version. In particular, the manuscript makes clear that, although crop models and land surface models were originally developed for different purposes and spatial scales, they are now both used for regional irrigation estimation, and that the aim of the study is to understand how differences in model structure and process representation translate into differences in irrigation and related fluxes. On this basis, I consider the revised manuscript acceptable in its present conceptual framing.

However, I would still like to express one residual concern. I am sorry, but I still do not understand why the study uses a generic crop with a fixed January 1 to August 31 cycle instead of adopting a more realistic agronomic choice for the Po Valley. Even if the purpose is not to reproduce local management in full detail, this crop calendar appears difficult to justify from an agronomic point of view. In particular, while the end of the cycle might perhaps be acceptable for some early-harvest systems, the choice of January 1 as the start of the crop cycle does not seem realistic for the dominant irrigated cropping systems of the study area.

For this reason, although I am willing to accept the revised manuscript, I encourage the authors to acknowledge this limitation more explicitly in the final version and to clarify that this crop setup should be interpreted only as part of the conceptual intercomparison framework, not as a realistic agronomic representation of the Po Valley. For these reasons, I recommend minor revision. The manuscript is now largely satisfactory, but the authors should clarify in the text that the January 1–August 31 generic crop cycle is a conceptual simplification adopted for intercomparison purposes and not an agronomically realistic crop calendar for the Po Valley.

We thank Marco Acutis for his second review of this manuscript. We now explicitly acknowledge this concern in the manuscript:

L551-554: *While the generic C3 crop is used to approximate the average behavior of C3 crops within a grid cell, it does not represent the full diversity of cropping systems in the Po Valley. In particular, major irrigated crops such as maize (C4) are not represented. Furthermore, the prescribed January–August crop cycle remains a conceptual simplification adopted for intercomparison rather than a realistic agronomic calendar for the region.*

We also clarify that the crop only starts growing when temperatures are warm enough:

L184-185: *Crop growth is limited by temperature and only occurs when temperatures exceed the base temperature of the generic C3 crop, set to 8 °C.*

Referee #4

Review #2 on « 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 »

I thank the authors for their point by point response. Most of my comments have been adressed but I still have the following two major concerns :

We thank the reviewer for reviewing the manuscript a second time. A response to the comments is given below.

- about different total available water used by the two models. The authors argue that they want to do the comparison « while leaving model-intrinsic structures and their parameters untouched ». I totally agree with this statement but the soil hydraulic characteristics are not model parameters to my opinion and it is not difficult to use the same pedo-transfert function for both models. In addition, they answer that, at the seasonal scale, water losses (transpiration) play a major role on irrigation amount prediction. I agree that transpiration is well known to be the main term of the water budget but it is highly dependent on the available water as well. I understand do not want to run the simulations again with the same TAW, but they need to, at least, to acknowledge that this is a limitation of the study that biases the comparison.

Simulations with harmonized soil hydraulic parameters were tested but led to physically inconsistent behavior in Noah-MP. We therefore retain the native parameterizations and clarify in the manuscript that differences in SHPs are considered an inherent part of the process-oriented intercomparison.

L525-532: *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. <u>Therefore, rather than harmonizing these parameters, we retain the native model parameterization and interpret the results within a process-oriented intercomparison framework, rather than as a direct comparison with identical parameter settings.</u>*

- about the comparison with irrigation satellite retrieval. To my previous comment « 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) » the authors argue that « Given the current length and level of detail of the manuscript [...] we have chosen not to extend the analysis further. » but I just suggest to add the estimates to existing figures. If the analysis is not extended further, I feel that it makes the paper more complex as many times in the manuscript. Indeed, the authors repeatedly state throughout the paper that satellite-based irrigation estimates are uncertain, even though no objective scientific evidence is provided to support this claim (the only use of these data is for comparison with the regional estimates in Figure 3). I thus propose to discard the use of this dataset to ease the understanding of the study.

We agree with the reviewer that the inclusion of this dataset without a more thorough analysis could be confusing. Given the uncertainties associated with the product and its limited contribution to the core objectives of this study, we have removed this comparison from the manuscript to improve clarity and focus.

Figure 3 and its description have been replaced to exclude the comparison with Dari et al. (2023) and now covers all the simulations years and sums the irrigation over the entire year:

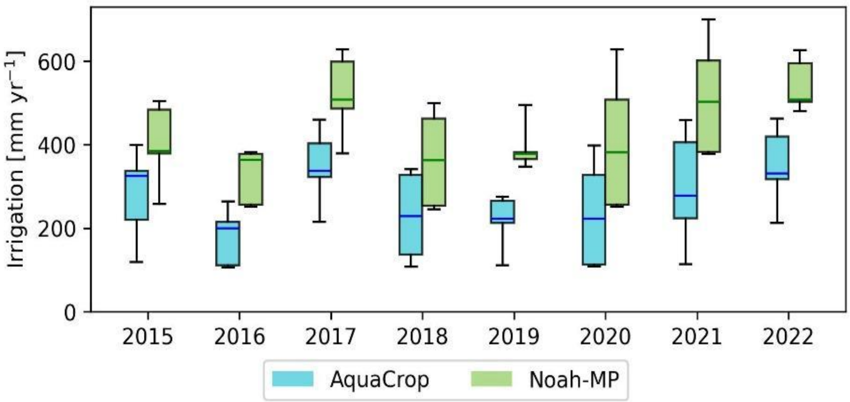

*Figure 3: Yearly irrigation amounts [mm yr^{-1}] for 2015 through 2022 for AquaCrop (blue) and Noah-MP (green). The boxes represent the irrigation values within the interquartile range (IQR), the lines in the boxes correspond to the median, and the whiskers extend to Q1-1.5*IQR and Q3+1.5*IQR or are cut off if all data points are within the interval (outliers are not shown).*

L360-369*: Figure 3 presents spatial boxplots of the annual irrigation rates for both models. More specifically, the interannual variability of the irrigation estimates is shown following the x-axis, and the spatial variability over the domain is represented by the extent of the boxes. First, as expected given the shared meteorological forcing, the temporal evolution of median irrigation is similar in both models, with Noah-MP consistently producing higher irrigation amounts, in line with the average annual irrigation rates shown in Figure 3. Second, the spatial variability also follows the same trends for both models, with a reduced variability in 2019 due to less variation in summer precipitation and temperatures over the domain. On average, 2017 and 2022 appear to be the years with the most intensive irrigation, followed by 2021, all years being very dry (Baronetti et al., 2020; Montanari et al., 2023). The results presented in our study are mainly driven by the meteorological forcings and have no limitation in irrigation water* *usage, while in reality, 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.*

The description of the dataset has also been removed, along with former Figure 4 and its associated text, to further improve the clarity and conciseness of the manuscript.