First, we would like to thank the reviewers for having carefully read the paper and provided valuable comments which helped us to improve the quality of the manuscript. We have taken into consideration all the comments raised by the reviewers and changed the manuscript accordingly. The details of our changes are highlighted in the main text. The point-by-point answers to Reviewers #1 and #2 are provided in the following and highlighted in red.

RC2: 'Comment on hess-2023-214', Anonymous Referee #2, 11 Oct 2023
The paper of Lombardi et al. deals with an interesting topic concerning the reliability and usability of satellite rainfall products for hydrological applications. To this aim, they consider a case study in northern Italy (Tanaro watershed) and test different precipitation fields achieved using only rain gauges, satellites, or several merging options. Precipitation fields are used as inputs for a hydrological model. According to the behaviour of the resulting hydrograph, the precipitation fields are judged in terms of reliability, supporting the discussion with several (very many, indeed) skill scores.

Despite the interest in the topic, I believe several changes in the paper are needed before it can be considered ready for publication in HESS.

First, I have concerns about the methodology followed. As a first step, since the main point is the reliability of the satellite rainfall products in providing quantitative precipitation estimates, I would have set up a validation exercise with observed (rain gauge) precipitation using, e.g., a leave-one-out or cross-validation approach.

> Thanks for the comment, it was intentionally decided not to use this approach, relying exclusively on an indirect method of validation through a hydrological evaluation. The work does not aim to do a local validation but a distributed one, thanks to the analysis of data relating to the drainage network.

Then, an indirect validation through simulated hydrographs should be justified more strongly, mainly if performed with a not-calibrated hydrological model (L202). This approach could be tricky and misleading because of the inherent limitations of the non-calibrated hydrological model so that errors can counterbalance and smooth each other. I suggest a preliminary calibration of the hydrological model with a reference precipitation field (e.g., only rain gauges) and only afterwards assess the changes caused to the modelled hydrograph with other methods.

> The reviewer is right; this sentence is misleading. It was meant that the model was calibrated in past studies on the Po basin of which the Tanaro River is a tributary (Coppola et al 2014) and not for this study. To avoid confusion, it has been deleted from the text.

Another concern is the description of the CA-based technique for interpolating and merging precipitation, which is unclear to me. An example should be given. Among other things, it is unclear why the authors also consider the time evolution of the precipitation field.

Regarding the technique adopted in detail, the review can refer to Coppola et al, 2010. Effectively, a detailed description of the CA algorithm is missing in the manuscript and the following section has been added.

The temporal evolution of the rainfall field is fundamental for the hydrological simulation, in particular, the CHyM model is forced with the hourly rainfall fields.

Last but not least, the paper could be structured much better.

The reviewer is right; the document has been rearranged.

The different options for achieving the precipitation fields are not clearly presented (e.g., some acronyms are provided in Fig. 2, then they are explained much later -L400- and not in the Methods, but in the Results section).

A description table has been added (table 2)

It is unclear why the authors decided to rely on 17 scores. This choice is somewhat confusing, in my opinion.

The 17 scores are necessary to obtain an objective evaluation of the analysis, given that each score highlights different characteristics. Obviously, if they are all positive, the results are confirmed as reliable.

Furthermore, many of the methods are presented in the results section or even in the conclusions (please refer to specific comments below). The discussion should refer to similar analyses performed by other authors to contextualise the results better. The conclusions section should be more than a summary of the paper.

In summary, I saw the possible added value that this paper can bring to the scientific community. Still, a thorough review is needed regarding the methodological approach and the structuring of the article. Please find below some other minor to moderate comments. I hope my review helps improve the quality of the paper.

1. *LL46-49: please revise. Dealing with predicted rainfall, it's impossible to remove uncertainty (correctly, in fact, the second sentence of the paragraph refers only to observed rainfall data). I suggest focusing on why accurate spatial distribution of rainfall observation is important.*

To avoid confusion, the sentence has been corrected as follows: “As far as the operational activity is concerned, the hydrological models are usually forced both with observed and forecasted rainfall data, and the uncertainty of hydrological forecasts is strongly related to the uncertainty of the input rain field. Therefore, providing hydrological models with observed precipitation data that is as realistic as possible becomes essential

in mitigating uncertainty, during the spin-up phase of the simulation when the hydrological model is forced with observed rainfall data."

2. *L104: "The work": do the authors refer to Shi et al. (2020)?*

Yes. The sentence has been corrected.

3. *LL111-112: I can't entirely agree with this statement. Indeed, there are a lot of studies dealing with this topic.*

Done. The sentence has been deleted.

4. *LL118-119 is a repetition of the second main objective declared at LL115-116. If the authors agree with my comment, please consider if the previous sentence (LL117-118) is well-placed and contextualised.*

The reviewer is correct; the second sentence LL117-118 has been deleted.

5. *L120: are all these 352 stations really useful? I guess the authors only need those lying into or close to the analysed watershed. From this point of view, it's unclear why the authors consider a much broader spatial domain than the investigated watershed (which, moreover, is not at the centre of the domain itself). I guess many stations, for example, lying in the north and northwest, are useless for this case study.*

The objective of this study is to validate a prospective operational framework designed for civil protection monitoring and forecasting. The rain gauge network employed in this research is the official national network. Utilizing all available data enhances the model's ability to effectively distribute the rainfall field, representing not a limitation but rather an added value. As rightly noted by the reviewer, the inclusion of 352 rain gauge data may be excessive, especially considering that only 73 of these data points fall within the specific basin under consideration. At the state of the art, a study aimed at understanding how much rain gauge data is sufficient to have the same results has not yet been carried out A possible future work could be to estimate the optimal number of rain gauges, based on their distribution.

It is crucial to acknowledge that the algorithm employed in this study applies smoothing to the processed data. Therefore, for scientific rigor, the entire operational domain is utilized, acknowledging both its advantages and limitations. One such limitation is the absence of rain gauge data beyond national borders, which could potentially impact the results. The primary focus of this work is to ascertain whether, despite these constraints, the model's performance improves when incorporating satellite data using this technique. Further investigations could delve into estimating the optimal number of rain gauges based on their distribution, representing a potential avenue for future research.

6. *L146: 1700 m3/s is a peak flow, average daily flow or what else? Some lines below the authors refer to a peak of 4350 m3/s (if "Autorità di Bacino del Fiume Po" is a reference, please add the year; if not, please explain/translate it in English).*

- A maximum flow discharge that can reach 1700 m^3/s in spring and autumn.
- The most significant of these events occurred in November 1994, when the entire river valley was damaged (Marchi et al., 1996; Luino, 2002) and the sensor at Montecastello, located at the outlet of the river recorded a maximum flow discharge peak of 4350 m^3/s (Autorità di Bacino del Fiume Po).

"Autorità di Bacino del Fiume Po" is the official Italian authority that releases information regarding the data relating to the various rivers, the document is among the references and was last viewed in November 2023. "Autorità di Bacino del Fiume Po" has been translated into English as suggested by the reviewer.

7. *L 165: Eq. (1) is quite ambiguous. It is well known that the reference area is hardly a radius, especially in orographically complex regions such as the Alps.*

This equation refers to Shi et al. Estimating a radius of influence is certainly not so immediate, but it can be a good starting point for making various considerations.

8. *L176: Earth*. Done. it has been corrected.

9. *Figure 2 and related caption: please revise. There are several errors: e.g. Guage uncal, "each case studies" [study], "eight [...] setting" [settings]. Furthermore, the terms uncal, cal, uncal1, uncal5, etc. are explained much later (L400). The explanation of the different inputs for the eight simulations should be highlighted much better (maybe with a devoted Table?).*

As suggested by the reviewer, table 2 summarizing the different inputs has been added.

10. *L325: finds*. Done. Now, LL335

11. *LL360-364: I guess this is methods, not results.*

    Section 5.1 also is not results, but data and methods.

    L406: it's 5.2

Thanks for the advice, this section has been reorganized.

12. LL418-419: however, while some of the stations considered in the study are located much further north of the watershed (as I claimed before, I believe they

are useless to this study), after the French border, there are no stations surrounding it. This drawback should be discussed.

The reviewer is right, as said before, this falls within the limits related to the selected domain, which we voluntarily wanted to maintain. In operational conditions we come across these types of problems, so the use of satellite data helps us overcome the problem, as demonstrated by the improving results.

13. LL519-528: that's methodology.

The reviewer is right, this is a repeat. the sentence has been deleted