Point-by-point reply to the comments

Anonymous referee #3

- **This paper presents a comprehensive modeling study examining the effectiveness of nature-based solutions (NbS) for mitigating both floods and agricultural droughts in Belgian catchments. The authors use a physically-based hydrological model (MIKE SHE/MIKE 1D) to evaluate various NbS measures and emphasize the importance of soil drainage characteristics in determining their effectiveness. While the study addresses an important and timely topic with a reasonable methodological approach, several significant concerns limit the strength of the conclusions. The main issues relate to model validation, oversimplified parameter representations, and methodological limitations in the comparative analysis between catchments. Despite these limitations, the work provides valuable insights into spatial targeting of NbS and contributes to the growing literature on nature-based flood and drought management. Detailed comments are provided below, which I hope will be useful in clarifying and strengthening the manuscript:**

Thank you for your review. Below, we provide answers to all your comments and explain how we have modified the manuscript accordingly.

- **Page 13, Line 295: I'm concerned about how the authors validated their saturated zone modeling. They mention doing a "visual assessment" against soil drainage classes (Table 3), but don't explain what this actually means or how they did it. Without any real groundwater measurements to compare against, they're basically checking if their model matches the same soil data they used to build it in the first place - which isn't very convincing validation.**
  Since the whole paper hinges on how soil drainage affects these nature-based solutions, this validation is concerning. The authors should either explain their validation approach much better or be more upfront about this major limitation.

We acknowledge that this validation approach is experimental and still open to improvement. However, we still think it is valuable as it offers the advantage of leveraging the limited data available in the study site to assess whether the simulated hydrological behavior of soils reasonably reflects field observations, and does so in a fully spatially distributed way. This remains uncommon in hydrological modelling, where validation often relies on a relatively sparsed point-based measurements (gauging stations, piezometers, sometimes soils moisture probes).

We would also like to clarify that the soil drainage class information from the Belgian soil map is mostly independent of the parameterization of soils in the model. It is used only for the spatial discretisation of soil units. The hydraulic properties assigned to each soil unit, including retention and conductivity curves, are not derived from the drainage classes. Instead, they are generated solely using the EUPTFv2 pedotransfer function, based on an independent map of soil textural fractions and the depth of the soil horizon as predictors.

Moreover, the modelled dynamics of the saturated zone are influenced not only by the soil prarametrization but also by other input parameters such as topography, geology, vegetation and climate. This validation therefore helps to verify that the combination of these parameters results in a reasonably coherent representation of soil functioning across the model domain.

In the manuscript, we have revised line 295 to clarify this approach and have added a direct reference to Figure 4, which presents the outcome of the validation.

- **Page 15, Table 4: The translation of nature-based solutions into model parameters seems overly simplistic. For instance, representing soil pitting as just adding 20mm of water storage ignores the actual geometry and connectivity of these features. Similarly, hedgerows create complex root systems and flow pathways that go well beyond modifying just the top 40cm of soil properties. The authors should elaborate on this.**

Yes, we agree with your point. It was also raised by the reviewer 2 in his initial review. You mentioned two examples, but there are likely many more; at this stage, evaluating the appropriateness of NbS parameterization remains largely based on expert judgement. However, we believe the advantage of our approach to parametrize NbS, compared to more conceptual modelling strategies, is that it is grounded in physical parameters, which are, for the most part, measurable. As the literature on NbS continues to grow, this opens the door to improved parameterization based on concrete field measurements rather than on calibration of model parameters. However given the current state of knowledge, we prefer consider our NbS parametrization as hypotheses (we recognized the uncertainty associated), and our results depend on these assumptions.

This is why we mentioned just above the table 4 (in lines 336 – 340) that : *"It should be noted that although some literature exists on how NbS may influence these parameters, significant uncertainties remain regarding their exact values, as they cannot be assumed to apply* *directly to this specific case. These parameter values should be considered as hypotheses, while remaining within plausible ranges."*

We also acknowledged this limitation in our section 3.7.2 (line 680) : *"Three major sources of uncertainty affect estimates of NbS effectiveness in our study: … (iii) the parameterization of NbS (Brauman et al., 2022) … The third source has not been explicitly quantified. Instead, we compare our results with reported NbS effectiveness in the literature to assess whether our findings align with expected trends. While this provides a useful reference, variations in study contexts may introduce further uncertainties. A sensitivity analysis of NbS parameterization, spanning a plausible range of parameter values, could offer valuable insights into how parameter input uncertainties influence model output uncertainties. Improving the robustness in the estimation of NbS effectiveness remains a critical area for future research."*

We have added a few words to this study limitations section emphasizing the need for more field-based evaluations of NbS effectiveness, under various conditions, which would permit direct comparisons with model outputs and provide validation of the model's ability to represent these measures accurately.

- **Page 19, Lines 415-419: The comparison between the two catchments has some fundamental issues that weaken the conclusions about soil drainage effects. The authors are comparing completely different systems - different geology, topography, precipitation rates (C2 gets 20% more rainfall), and entirely different types of NbS measures (hedgerows vs. peatland restoration). With so many variables changing at once, attributing the differences in effectiveness solely to soil drainage characteristics is problematic. A fair comparison could be testing the same measures across areas with different drainage within the same catchment.**

In this section, our intention was not to attribute the differences in overall scenario effectiveness to the natural soil drainage characteristics of the catchments. The purpose was simply to present the results of the two scenarios, which were implemented in different catchments, and developed in accordance with the predominant land use (agricultural or forested) by de "Schéma stratégique Vesdre". This section should be seen as a preliminary result intended to contextualise the rest of the results. It was not our aim to explain the origin of these differences here. As you rightly pointed out, this would require a more in-depth analysis.

We have added a sentence after line 433 stating that these results compare the overall outcomes of two different scenarios applied in two distinct catchments, which differ in land use, geology, topography, and precipitation rates. Therefore, these results should not be used to generalize the effectiveness of the specific measures implemented within each scenario.

- **Page 25, Figure 5: The spatial effectiveness analysis presents interesting findings, but there are some methodological considerations that could strengthen the conclusions. The study uses the Belgian soil drainage classification for both model validation and to demonstrate that soil drainage controls NbS effectiveness, which creates some circularity in the reasoning.**

We do not see a fundamental issue in using the Belgian soil drainage classification both for model validation and to illustrate that soil drainage influences NbS effectiveness, as the validation was carried out solely on the "baseline" scenario and the Belgian soil drainage classification was not used to calibrate any of the NbS measures.

- **The statistical analysis in Figure 7 would benefit from addressing the highly unbalanced sample sizes across drainage classes (n=15 to n=18,851), which may affect the reliability of the significance tests.**

Yes, this difference in sample size between drainage classes is explained by the uneven distribution of land use types across these classes. For example, in agricultural contexts, most croplands are established on well-drained soils, while waterlogged soils are generally avoided. This naturally leads to highly unbalanced sample sizes when analysing the effects of NBS by drainage class.

We adapted the significance tests by applying them to a subset of 200 data points randomly sampled. Significance tests results were removed for samples smaller than 200 data points. We also have corrected Figure 9 in the same way. Some significance tests results slightly changed but in general our previous observations and our reasoning holds, so we didn't made any change in the main text.

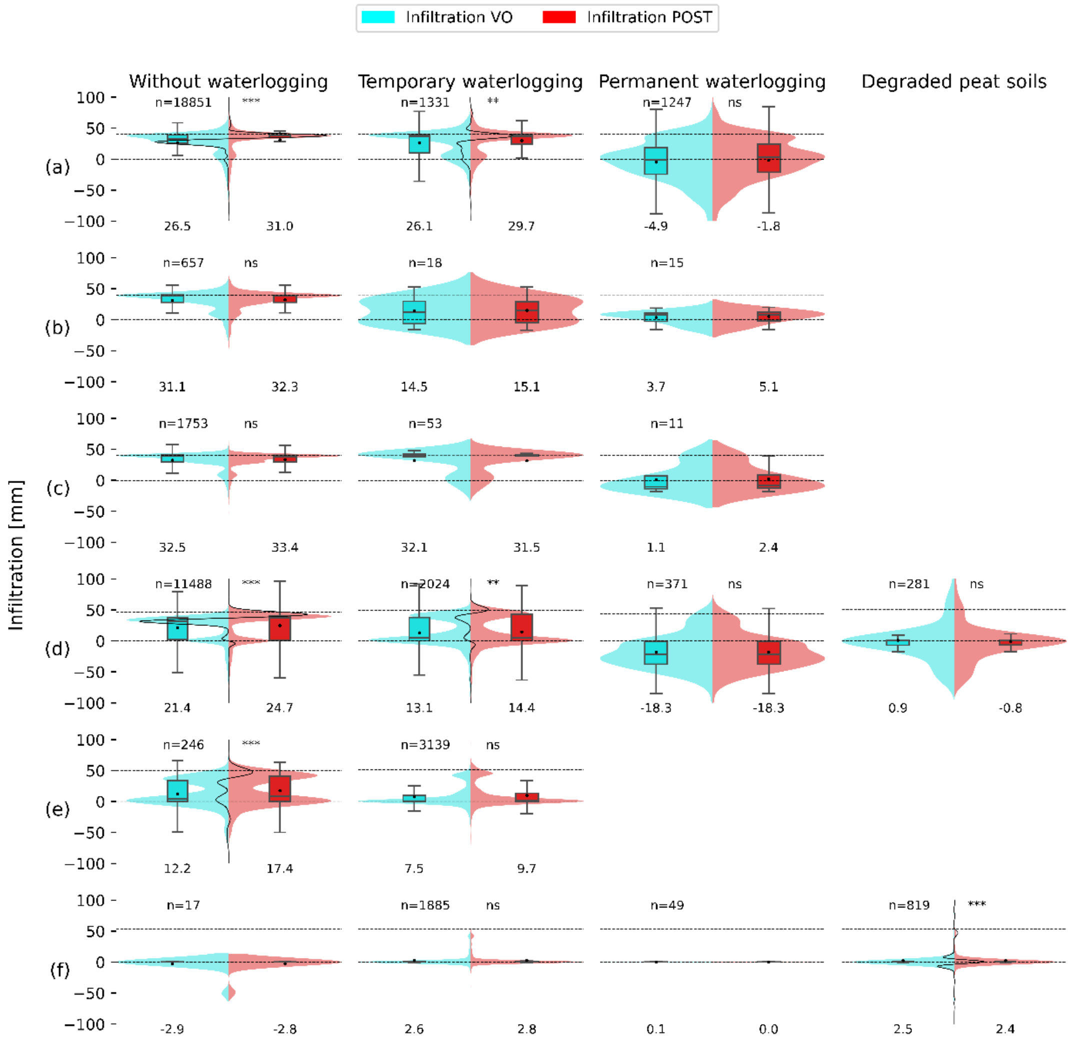
- **Additionally, since different NbS measures were implemented in different locations, it's challenging to isolate whether effectiveness differences stem from soil drainage characteristics or from the spatial distribution of the measures themselves.**

Yes, we agree with this point. At the very least, there appears to be a correlation between the improvement of infiltration and the soil drainage class where some measures are implemented. But is this truly due to the drainage class itself, or to other correlated factors, such as topography? It is clear that other factors also play a role, and we have clarified this point more explicitly in the manuscript.

That said, we are strongly convinced that this is not merely a correlation, but that there is indeed a causal link between soil drainage class and the effect of certain measures. This is what we aim to explore further in the following discussion (lines 540–560) and in Figure 8, by analysing some of the mechanisms that could explain this causal relationship.

- **The analysis focuses on one extreme event (July 2021), and it would be valuable to see how these patterns hold across different hydrological conditions. The authors acknowledge that their 20-40m model resolution may limit representation of fine-scale processes, which is particularly relevant for practical NbS placement decisions.**

This figure presents the same analysis as Figure 7, but for a different rainfall event (from 21 to 23 August 2007) with a total precipitation of approximately 50 mm (compared to approximately 200 - 250 mm for July 2021). Significance tests were also performed on a sample of 200 points. In our view, the observed patterns remain relatively consistent. To maintain coherence and avoid further lengthening an already dense manuscript, we prefer to leave this information here and not include it in the main text.

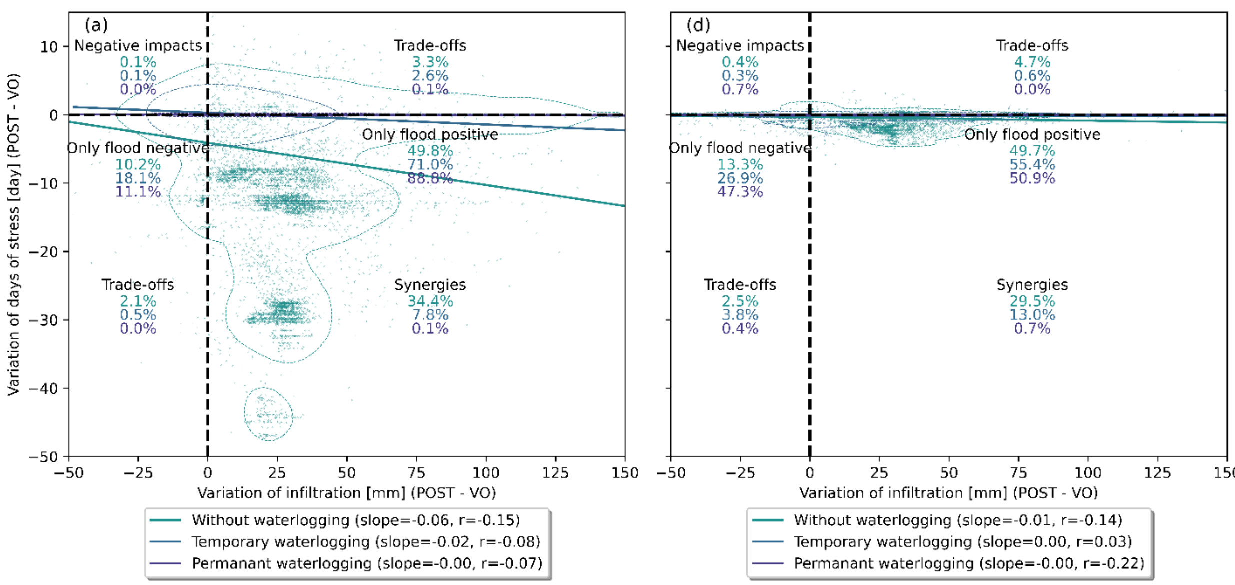


- **Page 30, Figure 10: The synergies and trade-offs analysis is interesting but feels somewhat predetermined by the study design. Since the authors specifically chose measures that enhance root systems (hedgerows, forest diversification) and placed them primarily in well-drained soils, it's not surprising that these same measures show benefits for both floods and droughts. The conclusion that root depth is the key mechanism linking both benefits makes sense conceptually, but the analysis doesn't really test this hypothesis. A more convincing analysis would compare the same measures across different soil types or test measures that specifically don't involve root system changes.**

We are aware that our approach is only one among others, and that it can be improved. In the manuscript, we already acknowledge the limitations of our approach, including uncertainties, and highlight that it opens the door to new perspectives and research directions.

Here (bellow) is the comparison you suggested: a figure similar to Figure 10, but focusing only on hedgerows (panel a, left), which involve changes in root depth in the modelled scenario, and forest practices aimed at limiting soil compaction (panel d, right), which do not involve changes to the rooting system. Both are tested across different natural soil drainage classes. These additional results appear to support our conclusions.

While we find this additional result interesting and consistent with our main findings, we feel it does not substantially enhance the value of the manuscript and primarily serves as further support for conclusions already well established in the main text. Given the manuscript is already quite dense, we prefer not to include it in the main body. We still remain open to including it as an annex if the editorial team deems it necessary.



- **The low percentage of areas showing drought benefits (27.3%) also raises questions about whether these measures are really as effective for drought mitigation as suggested.**

This low percentage is due to the fact that measures were not specifically implemented in what we defined as drought-prone areas, as it could be observed from the figure 9 on the distributions of days of drought (matric potential of -30m and return period of 25 years) before (BASELINE) implementing NBS. As a result, their potential to mitigate drought could not be fully expressed in locations that were not affected by drought conditions.

Referee #2: Martin Volk

- **Dear authors,**
  you have done a good job and the quality of the manuscript has improved considerably. You have considered and discussed all my suggestions and comments, and in an excellent way. The only thing that needs to be changed is the abstract. At over 550 words, it is far too long. It needs to be shortened. And keywords need to be added.
  Thank you and best wishes,
  Martin Volk

Thank you, Mr Martin Volk, for you review. We have shortened the abstract to approximately 320 words. Unfortunately, we could not find any specific guidelines regarding keywords in the HESS submission instructions, and did not find where to place them. If possible, we suggest to include the following keywords : Nature-based solutions, Floods, Agricultural drought, Hydrological modelling, Soil properties.