Point-by-point reply to the comments

Reviewer 1

- **The manuscript is engaging, well-written, and provides valuable insights into the effectiveness of Nature-Based Solutions (NBS) in the two selected case studies. However, I believe the authors should further clarify the differences in soil characteristics between the two case studies. This is critical since a key focus of the manuscript is on how soil properties influence the efficiency of NBS. Although the two case studies differ in geomorphologic and land use/land cover properties, they share similar soil characteristics (both predominantly silty). Emphasizing this aspect would help differentiate the results and strengthen the analysis of case-specific outcomes.**

Our intent in the manuscript was not to contrast the effectiveness of nature-based solutions (NbS) between the two catchments but rather to compare the effectiveness of different NbS (related to land use) across different pedological contexts (primarily based on natural drainage characteristics) that are present in both catchments, albeit in different proportions. However, we agree that a more detailed description of the soils in the study area would be beneficial.

To address this, we have added a dedicated section titled **Soil Context** in the **Materials and Methods (Study Area)** section, providing a more comprehensive description of the soil units. In this section, we describe the key geological and pedological processes that have shaped the current soil characteristics in the study sites. We also explain how the soils differ, with a particular focus on their natural drainage properties and spatial distribution.

- **Additionally, I suggest some improvements to the figures, as the distinction between the two case studies is not particularly clear in some of them (e.g., Figure 3 and Figure 5). Enhancing the visual representation could better convey the comparative analysis.**

For figure 3, we have added a stacked bar chart of the proportion of surface of natural drainage classes of soils found in the two case studies. This would help to differentiate the catchment in terms of soil natural drainage characteristics.

For figure 5, we have splitted the figure in two parts; one for each case study.

- **Lastly, lines 312–315 are particularly intriguing. The unexpected results warrant a deeper explanation. It would be beneficial if the authors elaborated on the**

potential reasons behind these findings, focusing on how geomorphological and soil characteristics of the selected areas may have influenced the outcomes. Given the proximity of the two areas, climatic factors are likely not a significant contributor, so discussing the impact of local geomorphology and soil properties in greater detail would add valuable context.

To avoid any ambiguity, we must clarify that lines 312–315 refer to the peak discharges modelled for the BASELINE scenario (without NBS implementation). This clarification have been be added to the manuscript.

The paragraph primarily aims to recontextualize the two catchments by explaining that, in the current (baseline) scenario, the forest-dominated catchment (C2) generates more runoff than the agriculturally dominated catchment (C1). In the manuscript, we use the term "counterintuitive" to describe the observation that the C2 generates more runoff than the agriculturally dominated catchment C1. However, "counterintuitive" might be too strong a term, as these results are entirely expected and align with our understanding of hydrological functioning of both catchments. Observations at the monitoring stations further confirm this, so it is not surprising. Our intention was simply to highlight that this does not align with the common assumption that forest-dominated catchments produce less runoff than agriculturally dominated ones.

We have addepted the paragraph in order to clarify the interpretation.

Also we have replace the term "climate" with "precipitation rate", which is more valid.

We have also added a sentence about the impact of local soil properties on the modeled hydrological functioning of the catchments, referring to the "soil context" section in the Materials and Methods, and the geological and soil maps figure.

Since this paragraph is intended solely to recontextualize the two catchments by referring exclusively to the BASELINE scenario—and given that the primary objective of the article is not to compare their hydrological functioning under this scenario but rather to assess the effectiveness of NBS—we prefer to keep it concise and maintain focus on the article's main message.

Reviewer 2

- **Abstract**
  Well written. I would already here mention what kind of model you used and what

NbS you investigated, makes it more specfic. If you cannot mention all, give the most important ones / examples.

We have added a small description of the model we used and also mentionned the investigated NbS.

- **I think you could also highlight a bit more that is an experimental study (in my opinion), because you also investigate in how far the model is able to simulate the effect of the retention measures. I would write this.**

We highlighted this point by mentionning that : "This study presents an innovative approach to modelling the effects of NbS landscape planning scenarios, explicitly simulating soil water fluxes. This approach allows for the investigation of how the spatial variability of soil properties influences NbS effectiveness in mitigating both floods and agricultural droughts."

And that : "The modelling approach was validated by accurately reproducing river discharge and saturated zone dynamics. It effectively captures soil natural drainage characteristics and provides a reasonable representation of NbS effectiveness, as indicated by consistency between simulated and literature values."

- **Point out what is innovative, what was surprising, and what are pros, cons and limits of your general and modelling approach. It should made be clearer what the value of your study is.**

We have modified the end of the abstract, covering these aspects : "Results highlight that the evaluation of NbS effectiveness should recognize the spatial variability in their performance. This variability should inform the type and location of NbS to increase their overall effectiveness. The study underscores the need to move away from siloed evaluations of NbS and instead adopt a coherent, territory-based approach. Our study may serve as a basis for discussion and action, supporting decision-makers in implementing measures to enhance hydrological resilience. This modelling approach provides a solid foundation with potential lines of futur research to refine NbS effectiveness assessments, such as strengthening data availability and working on uncertainty analysis."

- **1.Introduction**

  General comments: Well elaborated. For me, however, it is not clear whether the two study areas were already affected by floods and droughts, so perhaps this could be added here (also in the description of the study areas). I hope I did not miss such a statement in the manuscript.

We have added a paragraph on this in in the first part of the study area description, focussing on what was the motivation to choose this area specifically.

- **Line 47 ff (NbS): I think that the wealth of different "perspectives" and opinions on NbS can be confusing, you can be a bit more specific regarding your measures (some of them seems to be Natural (Small) Water Retention Measures ;o)). Have a look at our paper that tries to bring a bit more order into all these concepts:https://www.mdpi.com/2071-1050/16/3/1308. I was thinking of it when reading the concerning sentences in the manuscript. Just have a look at our paper, if it helps, that's fine, if not, that's fine too. No obligation / need to cite, only if it helps.**

We have added a mention on other overarching concept, including NSWRM, in the introduction and we have cited your work, which clearly defines these terms. This would give readers a reference for further explanation of these terminologies.

- **Line 76 ff (models): You might provide a brief concise overview on some of such models either here or in the model framework description (2.2) which also lead to your model selection.**

Yes, we have added this paragraph in the introduction to adress you comment: “Several hydrological models have been used to simulate the effectiveness of NbS, including, among others, SWAT+ (European Commission, 2020), LISFLOOD (Institute for Environment and Sustainability : Joint Research Centre et al., 2012), HEC-HMS (Guido et al., 2023), or MIKE SHE (Fennell et al., 2023b). These models generally integrate different modules to calculate different water fluxes (infiltration, runoff, evapotranspiration) and can be based on empirical, conceptual, or physically-based approaches. For example, SWAT and HEC-HMS use the well-known Soil Conservation Service (SCS) method to calculate water infiltration into the soil based an empirical formula and the soil curve number (CN). Other models, such as MIKE SHE, are process-based and use physical equations, such as Richard's equation, to model infiltration. Some models allow for fully distributed simulations, while others are lumped models or adopt a non-distributed approach. Each model has its own advantages and limitations, and the choice depends on the specific needs of the study in terms of modeling capabilities, as well as the required spatial and temporal scale and resolution, and the available data (Kumar et al., 2021b).”

- **For me, it is in general still a problem that the measures we simulate in the model show an effect almost immediately, which does not correspond to reality. Depending on the landscape conditions, there is a delay in these effects (such as**

retention). If you see this similar, you could discuss that somewhere (intro or discussion, maybe better there).

Yes we see it similar, even though the main focus of the artcile is more on the spacial variability of NbS effectiveness, the temporal variability deserves also to be considered.

To do so, we have added in the section "**2.2.2 Landscape planning scenarios**" :

"The scenarios thus compare the current situation with a long-term projection (horizon 2050), without representing the transition process. It should be emphasized that accounting for the transition phase may be essential, as some measures require several years or even decades to achieve full efficiency, and the speed at which a measure becomes effective may serve as an important criterion for prioritizing NbS."

We also discuss this aspect in the "**3.7 Study limitations and knowledge gaps**" section :

"Furthermore, the effectiveness of NbS may vary over time, depending on (i) meteorological variability (e.g., antecedent moisture conditions) and (ii) the time required for a measure to become fully effective. For instance, some measures, such as forest diversification, may take several years or even decades to reach full maturity. In the current study, the transition phase between the BASELINE and POST scenario was omitted. However, time-variable effects of NbS may be important to consider when defining priorities in a specific local catchment context (Fennell et al., 2023b). This time-variable effects also raises the question of incorporating future climate scenarios into the modeling of NbS effectiveness (Gómez Martín et al., 2021)."

- **Line 86 ff (objectives) For me, as mentioned, it is also an experimental study that tests the capabiities of your (modelling) procedure for your task. I mean you clearly show the pros and cons and gaps (what still has to be done).**

Yes, we have added an objective on that :

"The paper addresses four key objectives: i) Building a modelling approach that represents both the spatial variation of soil charateristics (focussing on soil natural drainage) and spatialized NbS scenarios, as well as their interactions…"

And we have added in results that :

"Overall, the model successfully represents the variability in the natural drainage characteristics of soils in the BASELINE scenario. This addressed our first objective, which was a prerequisite for assessing the effectiveness of the NbS measures on various soils with differing natural drainage characteristics."

- **2.Materials and Methods**
  2.1 Study area
  General comment: See my comment before, add (maybe here) info if the study areas already faced floods and / or droughts.

See above.

- **Figure 1: The soil map is a bit hard to read.**

Yes, we have separated the soil map from the figure 1, to make it bigger and we have changed the symbology to make it clearer.

- **Line 106: Sentence "Apart from peatlands,.." . I think you can delete this sentence, it is already written in line 100 ("Soils are mostly silty.")(except of the stone content) and does not provide additional information).**

This has been removed. And we have rewrited a dedicated section in the Material and Methods on soil description.

- **2.2 Modelling framework**

  General comment: See my comment in the intro - provide a brief concise description on other models with similar capabilities (or weaknesses) such as MIKE SHE / MIKE 1D.

We have added a paragraph in the introduction that provides an overview of other hydrological models used to simulate NbS. To avoid excessive repetition, we do not repeat this information here. But only focus on MIKE SHE / Mike 1D capabilities, and how it is addapted to our task.

- **2.2.1 Hydrological model**

  Line 124 ff (data description): I would add a table listing the data and describe the most relevant characterists / information it. See this example (section 2.3 Model inputs)https://www.sciencedirect.com/science/article/pii/S1470160X1931012X?via%3Dihub

We have included a table that summarizes input data.

- **Line 137 ff: What about land management (tillage crop rotation, fertilization, etc.) data, for instance from agricultural statistics, data from Integrated Administration and Control System (IACS) or interviews?**

To our knowledge, systematic statistics on agricultural practices (tillage or fertilization, …) are almost non-existent or not publicly available in our study areas. In Wallonia, we have a public database that is anonymized and updated annually based on farmers' declarations, providing information only on the main crop of the current year. In our study areas, apart from permanent grasslands, most main crops were maize. The LULC map we used incorporates data from this database (2018), which enables us to distinguish different types of crops. Therefore, we decided to simplify croplands by considering two categories: Open production surfaces and croplands assuming maize as a representative crop for parameterizing vegetation development (LAI, Kc and root depth) in crops.

In the article, we have added : “Apart from permanent grasslands, which were considered open production surfaces, most croplands were cultivated with maize. Consequently, maize development dynamics were assigned to all croplands, and in the absence of information on cropping practices, no crop rotations were modelled.”

- **Line 201 (Moriasi et al., 2007): I know Daniel Moriasi and I highly appreciate his work, and I know that many people use this performance guideline, but I think it is not always the best method to evaluate model performance. In some cases it might not say very much about how well the hydrological dynamics are represented by the model (you describe the dynamics later in the text, so all fine (just a comment)).**

We agree with your point. The Moriasi guidelines are widely used, and we included them as a reference to provide readers with a familiar benchmark. However, we already acknowledge their limitations and discuss them in the text (line 534 – 540 on the first preprint).

- **2.2.2 Landscape planning scenarios**
  General comment: Are these scenarios and suggestions just your ideas or are they based on some River basin or landscape management plans or something like that in your region that suggest some of them?
  Would be good to know.

We have added in the beginning of the section “Landscape planning scenarios”:

“These landscape planning POST scenarios were developed in accordance with the “Schéma Stratégique Vesdre” (Inondations - Reconstruction, 2024), which outlines a long-term vision for sustainable territorial development and proposes a range of potential measures to mitigate flood and drought risks. The modelled scenarios (POST) incorporate as

many feasible measures as possible within specific contexts (agriculture-dominated and forest-dominated) to assess their potential for flood and drought mitigation."

The "Schéma Stratégique Vesdre" is also presented at the beging of the presentation of the study area.

- **Line 229 ff (Agricutural practices): These are partly very small areas (if I understand this correctly). Is it relevant?**

Yes, these crop areas cover only 2.6% of the total catchment 1 surface. Consequently, their impact on the hydrograph at the catchment scale (or outlet) is probably limited due to their small extent.

We should, however, mention that the agricultural practice we implemented (soil pitting for maize crops) is an incremental rather than a transformational measure. Unlike no-till farming, it does not require farmers to change their agricultural system.

Our scenarios were aimed at implementing as many feasible measures as possible in our study site to demonstrate their potential combined maximal effecteveness while staying true to the catchment's existing conditions (LULC). We sought to demonstrate a range of options, which we also evaluate directly where they are implemented. This is one reason we have chosen a fully distributed hydrological model: it allows us to assess the effectiveness of measures not only at the outlet but also within the watershed, directly where they are implemented and effective.

We did not make any changes to that point.

- **Table 3: The table is titled "summary of hypotheses.." - are there other papers that used these or similar parameter modifications to "describe" such measures? How to confirm or reject the hypotheses? By measurements? These are these experiments?**
  I guess there is no kind of a parameter database for such measures in the model (similar to SWAT) that can be extended?

We have added a disclaimer statemement about this limitation of our approach in the material and methods :

"These NbS measures are incorporated into the model by adjusting some key parameter values (**Error! Reference source not found.**). 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 discuss this limitation in the “3.7 Study limitations and gaps / Uncertainties” section :

“Three major sources of uncertainty affect estimates of NbS effectiveness in our study” … “(iii) the parameterization of NbS” … “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 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.”

- **2.2.3 Integrated hydrological analysis of model outputs**
  Line 248 (Hydrological indicators): So, you developed these indicators? I think there several indicators out like this, I think you should provide an overview. Infiltration and agricultural drought, etc., are not new "indicators". I would better point out what is new here (because - again - you write that you developed them) and what is better compared to the exiting ones. Do you see it as as a kind of "enhancement" to other hydrological indicators, such as the "Indicators for Hydrological Alteration (IHA)", introduced mainly by Richter et al. and Poff et al. and many more?

Yes, the names we assigned to the indicators we calculated or developed may not have been well chosen. We have renames them. We also have clarified which indicators were simply calculated (already existing) and which were newly developed.

- **3.Results and discussion**

  General comment: Well elaborated and written.
  3.4 Spatial variability of effectiveness of NbS against flood

  Line 426 ff (These findings..). The question is again) if your results - considering all uncertainties - justify such a statement (second part of the sentence)?

We have addapted this sentence. We also reviewed the manuscript to formulate more nuanced recommendations given all model uncertainties.

- **What about the delay of measures effects?**

See above.

- **Regarding flood protection measures - when are floods are too strong, that measures have no effect anymore?**

This paragraph concerns results of effectiveness of measures in term amount of infiltrated water focussing on the rainfall event between 13 and 18 July 2021 which was the most severe ever recorded in the region, with an estimated return period of 300–400 years. These results are therefore relevant to extreme flood events.

- **Line 468 ff ("These areas, we believe.." ff): See my comments above. Are the results reliable enough to state this - would you tell that a stakeholder / water manager, etc.? OK, you eaken your statement with last sentence, but think of reformulate it.**

We have addapted this sentence.

- **3.5 Spatial variability of effectiveness of NbS against agricultural drought General comment: You could perform allocation change experiments" (searching for the most effective measures (or measure combinations) at specific locations)? We use multi-criteria optimization for that (see https://www.optain.eu; see also project deliverables)). Could be a point for discussion?**
  3.6 Synergies and trade-offs between flood and drought mitigation with NbS General comment: See my previous comment regarding "allocation experiments" / combination of measures.

We have extended our last paragraph in section 3.6, discussing multi-criteria optimization :

“This opens the door to allocation change experiments and multi-objective optimization. Since different NbS or combinations of NbS can create both benefits and trade-offs across multiple (often conflicting) objectives, it is essential to identify one or several optimal implementation strategies that maximize benefits while minimizing trade-offs. Among other, one approach is based on Pareto-optimality (Deb and Jain, 2014), where a set of optimal solutions is generated such that no objective can be improved without worsening another. From this set of optimal alternatives, decision-makers can explore and the select appropriate solutions based on their priorities. Among the available tools for such analyses, the Constrained Multi-Objective Optimization of Land Use Allocation (CoMOLA) software (Strauch et al., 2019) has been developed to facilitate these assessments and has been applied, notably, within the OPTAIN project (European Commission, 2020).”

- **3.7 Study limitations and knowledge gaps**

  General comments: In general, I would emphasize the value of the study a little better.
  So far, the gaps are highlighted quite well in this chapter (although the word "however" is used a little too often ;o)), but it should be worked out a little better. What about the uncertainties (capability of the model(s) (what are the system limits (scale, describing the efficincy measure under the given circumstances, data, etc.? Can they (uncetainties) be quantified? What does that do to the reliability of the results? Is it sufficient so far to be able to derive reliable recommendations?

We have completely revised Section 3.7 to make it more structured and readable. We have organised it into five subsections:

- Data requirements
- Uncertainties
- Multi-scale assessment of NbS
- Multi-objective assessment of NbS
- Bridging the gap between simulated scenarios and real-world implementation

The comments you made regarding uncertainties are mainly addressed in the "Uncertainties" sub-section.

We have also rewritten this 3.7 section to make it more positive, highlighting through various examples the value of our study, outlining what remains to be done, and suggesting future research directions.

- **For me, this is a very good modeling experiment that could be used as a basis for further discussion (scientists, stakeholders), also to specify what needs to be tackled urgently next. To do this, you could involve farmers, authorities, economists, and first ask them for their opinion on the methods, measures and results.**

We tried to address this comment by adding a new subsection devoted to "**3.7.5 Bridging the gap between simulated scenarios and real-world implementation".**

- **I have already mentioned other points (taking into account the delayed effect of the implemented measures, experiments on the spatial allocation of measures (and the combination of measures), flood protection measures – when are floods**

too strong to be mitigated by these measures? etc.). You basically already have many of the points in this chapter, but I would discuss it a bit more clearly and in a more structured way and – as I said – emphasize the value of the study more clearly.

See our responses above.