Dear Dr. Sakic Trogrlic,

We would like to express our gratitude to the reviewer for the detailed feedback on our manuscript, and to you for leading the review process. We carefully addressed the feedback. Please find below the review comments (in black) and replies (in blue).

With kind regards, and on behalf of all co-authors,

Tim Busker

Reviewer #1

The authors have thoroughly addressed the main concerns raised in the previous review. The revised article has been significantly improved in terms of methodological transparency, analytical depth, and clarity of the comparative analysis between the Flood Forecasting and Early Warning Systems (FFEWSs) in the transboundary Meuse and Rhine river basins in Northwestern Europe. The authors have now reported a more systematic classification and synthesis of several important aspects of the FFEWSs, offering more insights into how and why the FFEWSs differ between countries, reflecting differences in forecasting methods (including in models ownership), crisis management procedures, and governance systems (federal vs. centralized mandated agencies). The revised summary table (Table 1) is now far more comprehensive and clearly structured, allowing the readers to find more clearly the key information retrieved by the authors for the different regions. This synthesis supports the comparative analysis of the key aspects of FFEWS across countries and regions that the paper aims to provide. Moreover, the authors added some useful information and insights on the technical improvements of the FFEWSs after the 2021 floods, on the remaining challenges and barriers (e.g., to develop impact-based forecasts), enriching the discussion as suggested. Finally, four insightful recommendations are provided and linked to the main limitations of the FFEWS. However, the article requires several further minor revisions to enhance its quality, readability and clarity. In particular, here below a summary of the main points is reported followed by detailed comments.

In the introduction, the two research questions exposed (Section 1.3, Research gap and aim) are too narrow with respect to what the paper actually achieves. They are valid but incomplete, as they do not anticipate the analysis of all the main operational and institutional characteristics of the FFEWSs, as they limit the focus on warning levels, which is only one of the points analysed in detail by the authors. Other findings are not introduced, including how these systems have evolved in response to the July 2021 flood event and which technical limitations remain (e.g., lack of impact-based forecasts). These aspects could be introduced as a third question addressed, to improve alignment between the paper's introductory framing and its findings.

The authors should reduce overlaps and repetitions between the Introduction and the Case study section (see detailed comments below).

On the methodological side, the interview methodology has been clarified (see Section 2.1), including how the interviewed experts were selected, and their role. The literature review approach is better contextualized, but is still not very transparent in terms of the search procedure and selected materials,

neither providing the full set of strings for the search, nor stating explicitly the number of documents found (only 16 are mentioned, and the reader is left to guess if they are all - probably not). The authors should further clarify these points for a fully systematic, reproducible literature search.

To summarize, significant improvements have been made, but further minor revisions are required to make the manuscript suitable for publication, enhancing readability and quality (see detailed comments below). These issues should be fixed through a full revision and proof-editing. If all the recommended minor revisions are addressed, the manuscript will become suitable for publication without further review needed.

We want to thank you sincerely for the detailed and valuable review you provided, and for the positive words about the changes we made. Based on your suggestions (above and below), we revised the manuscript. We believe this led to an improvement in the readability, clarity, and overall quality of the manuscript.

In the comments below, line (L.) numbers refer to the revised manuscript with tracked changes.

—----------—----------—----------—----------—----------—----------—----------—

Minor and technical comments

-- Main comments --

- Section 1.3 - Introduction: The two research questions are valid and reflect some in-depth analyses conducted, but are too narrow relative to the broader scope of the paper. The two questions emphasize warning levels, but do not fully capture the comprehensive analysis of operational and institutional characteristics of FFEWSs across the transboundary basins that the authors conducted. Additionally, the paper explores how these systems have evolved after the July 2021 flood and identifies remaining technical limitations, none of which are reflected in the current framing of questions dealt with in the introduction. Including one (or two) additional research question(s) to address these aspects would improve alignment between the study's aims and its actual contributions.

We agree that our research questions did not entirely capture the main results of the paper, especially after the revisions had been made. Moreover, there was also an overlap between the research goals and the research questions. Therefore, we deleted the research questions and synthesized them, together with the aspects you mentioned above, in three main goals:

> *"We intend (a) to compare the institutional and operational characteristics of FFEWSs (e.g. differences in warning levels across countries and communication protocols), b) to outline and compare the changes made after the 2021 flood event and (c) to develop recommendations to further*

improve the systems. This will provide a comprehensive overview of the state of FFEWSs and pathways towards translating forecasts to more effective early action."

- Section 2.1 - Case study region and approach: there are some overlaps and references between these section and Section 1.2 (Current challenges in FFEWS) in the Introduction: I think that the challenges in FFEWS in the Introduction (Section 1.2) should be presented in a more synthetic and general way, without details on the 2021 flood event (including early warning challenges) which would better fit Section 2.1. This would help avoiding repetitions (E.g., "In the Vesdre, Ahr and Geul catchments, a significant share of the local population did not receive warnings (see Introduction).")

Thank you for identifying the overlaps between those two sections. We agree that the event-specific information in Section 1.2 better fits in the case study section (Section 2.1). Therefore, we rewrote Section 1.2 to focus on the general challenges of FFEWSs and rewrote the case study section so that it outlines the main challenges observed during the 2021 event specifically.

- Section 2.2 - Approach: The "combinations of keywords" used for the literature search and mentioned here (Lines 205-207) should be reported explicitly (at least as an Appendix or Supplementary information) for full transparency of the methods. Also, the number of resulting documents from this literature search seems to be not reported or it is unclear if it's only the 16 selected articles mentioned in the following sentence ("for an in-depth analysis"). The number of official reports analysed should be clarified too.

Most of the results in the paper are generated using information from the 14 semi-structured interviews, and the scientific articles (n=16) and country-specific reports (n=14) that they forwarded. To supplement this information, a *non-systematic literature review* was done. This resulted in an extra 40 peer-reviewed scientific papers and 5 international reports. This was indeed not clear from the text in Section 2.2. The information with the number of documents has now been added.

It is important to note that the *non-systematic literature review* is used only to supplement the information obtained through the interviews. A *systematic literature review* would of course require full transparency in the literature review process. However, this paper is not intended as a literature review paper but rather summarizing expertise and experiences in Northwestern Europe regarding FEWS. To make this more clear, we changed the description in Section 2.2 to show that we use the literature review only to supplement insights (and reports) gathered through the interviews. In the revised version of our paper, we now also include more information on the type of articles and reports selected through the non-systematic literature review: 5 reports from international organisations (e.g. The Red Cross and World Meteorological Organization) and 40 peer-reviewed scientific papers.

- Section 3.1 and Table 1: Differences in lead times of hydrological forecasts are reported in Table 1 but seem to be not commented on in the text. It would be good to highlight these differences and if available to provide insights on the origin or reasons behind such differences. A spontaneous question arises: Why do forecast lead times differ so much (from 1-2 days in Luxembourg to 15 days in The Netherlands)? Are these large differences somehow linked to catchment sizes?

Thank you for noting that the lead time differences were not mentioned in the text. We now included this in Section 3.1. However, we have no information on the reasons for these differences. We agree with your hypothesis that it could be related to catchment size, but also other factors play a role, such as the meteorological forecasts used. For example, in Rhineland-Palatinate (RLP), ICON-D2-EPS meteorological forecasts are used, which have a lead time of only 48 hours (see Section 3.3). Therefore, we will not elaborate on this in the manuscript.

- Section 3: The spatial resolution of the operational hydrological models would be a relevant piece of information but is not reported in the text (nor in Table 1). Can the authors briefly report if there is any information available on this? It would be interesting to see if there are differences among countries also on that aspect.

We agree that this would be an interesting characteristic to show. Unfortunately, this is impossible as we do not have this information for the different systems. The spatial resolution of the models is not fixed but often varies (in space and time) and is user-defined (as stated, for example, for the LARSIM model, by Haag and Luce, 2007). We did not derive this information from the interviews, and therefore, we cannot systematically compare the spatial resolution of the different models.

We agree that it is important to emphasize that the same model type (e.g. LARSIM) can have a different model setup. Therefore, we inserted the following in Section 3.1:

> *"Even if the same hydrological model is used, important parameters in the model setup (e.g. spatial resolution; Haag and Luce, 2007) are often user-dependent and can still differ between the different countries."*

- Sections 3 and 4: Several statements describing the current issues of FFEWSs (post-2021 floods), needs and wishes of stakeholders and authorities would benefit from more explicit sourcing of information (a more direct reference or quotes). To enhance transparency and strengthen the qualitative grounding of the analysis, the authors should consider including anonymized quotes from interviewees or at least indicate whether insights stem from interviews (with which authority) or literature (which report or article). For instance, in Section 3.5 (Hydrological Forecasts – The Netherlands, lines 519–520), the following claim would carry greater weight if supported by a direct quote or attributed source: "Predictions are deterministic, although a strong need and wish exist to deploy probabilistic forecasting approaches". Similarly, in Section 4, where issues raised by interviewees are summarized in Table 2, brief excerpts from interviews could enrich the discussion and bring practitioner voices more directly into the analysis.

Thank you for this suggestion. We think that newly inserted quotes in the manuscript should be checked and accepted by the interviewees, even if they are anonymous. We prefer to keep the sentences in their original form and to not change them to quotes, as the current text is carefully checked and corrected by the interviewees. Sometimes, this even included slight corrections from what they said during the interviews. However, we agree with you that it is important to be more explicit about the information source. On multiple locations, we now mention explicitly that information stems from the interview. For instance, the sentence in Section 3.5 we changed to:

"Predictions are deterministic, although the Waterschap Limburg stressed in the interview that there is a strong need and desire to develop operational probabilistic forecasting approaches."

And in Section 4, we inserted:

- *"Our interviewees in RLP (Germany) explained that the hydrological thresholds based on return periods (see Figure 3) do not always reflect corresponding impacts, and that impact-based thresholds would be more valuable to decision-makers."*
- *"According to our interviewees in the region, this has eroded trust between different stakeholders, which is detrimental to the functioning of an entire warning system (Seebauer and Babsicky, 2017)."*
- *"However, our interviewees in Flanders pointed out that flood risk management is not a core business of De Vlaamse Waterweg authority, and therefore, they lack specific expertise on flood risk."*

Detailed technical comments, presentation quality, and English language issues

- The authors should use consistently either the present tense or the past simple and not alternate between the two as done now in most sections (especially in the Abstract and Results). For example, in the abstract: "Expert interviews across the region reveal that …"; "The assessment of warning systems showed …"; "The interviews also revealed …". I would suggest sticking to the present when talking about what the interviews and analysis show.

We changed the use of tenses in the whole manuscript, focussing on the abstract and results. Moreover, the manuscript was read and checked by native English language speakers to correct grammatical (and some spelling) errors.

- Abstract: I would suggest briefly reporting in the abstract all the four main recommendations and avenues for further research (those outlined in Section 5.2), as only one out of four does not seem the most appropriate summary of the findings.

This is a very good suggestion. We now added all 4 recommendations to the abstract, in a shortened form (compared to Section 5.2).

- Table 1: The caption should clarify that the information reported is relative to the time of the authors' analysis (present day), i.e., after the 2021 flood event. Otherwise, given the focus of the paper, it can lead to misunderstanding on whether it reports the state of the systems during the 2021 events. Also, the authors should check and clarify the text in the first column (5th row): "Main hydrological and model(s) …"; probably there is a typo (the word 'and' should be removed).

This is indeed an important detail to add. We changed the caption of Table 1 to:

> *"Characteristics of Flood Forecasting and Early Warning Systems (FFEWS) in different European regions and countries, reflecting the state of the systems at the time of this study, after the 2021 flood event."*

We also changed the typo in the 5th row of the table, by deleting "and".

- Section 1 - Introduction: check the use of verb tenses; at the moment, the authors often move from present to past tense in an inconsistent way, even within the same section; for example, at Lines 67-69 when describing general concepts.: "For example, uncertainty in the collection and processing of meteorological data (Fig. 1, Left) may lead to a 'missed' forecast event, where thresholds in the system were not surpassed, while observed water levels reached extreme heights … By contrast, uncertainties can also lead to so-called false alarms, where the predictions suggest that warning thresholds may be exceeded, while effectively they are not.". Here I would suggest consistently using the present tense everywhere.

We checked the whole introduction on verb tenses, but found no other change needed besides the one that you correctly raised. We changed this part of the introduction to:

> *"For example, uncertainty in the collection and processing of meteorological data (Fig. 1, Left) may lead to a 'missed' event, where thresholds in the system are not exceeded, while observed water levels reach extreme heights (Cloke and Pappenberger, 2009). By contrast, uncertainties can also lead to so-called false alarms, where predictions suggest that warning thresholds may be exceeded, while in reality they are not."*

- Section 1 - Introduction (Line 44): the authors should revise the following sentence: "The 2021 flood further showcased the potential of early warning systems". I do not think that this sentence (especially the wording "showcased the potential") is in line with the analysis provided by the authors and with the literature (e.g., Da Costa et al., 2026). The 2021 flood event seems to expose more significantly the shortcomings in the early warning system chains rather than their potential, given the deadly outcome of the flood event and the fact that the forecasts and early warnings did not lead to timely and effective anticipatory actions like evacuation orders. Recent additional literature on this should be cited, e.g., Da Costa et al. (2026).

Thank you for this important comment. We completely agree with you, and therefore rephrased and expanded the sentence to:

> *"However, the 2021 flood revealed significant shortcomings in the early warning system chains. While exceptionally extreme precipitation was predicted by weather models (Busker et al., 2025; Mohr et al., 2023), and warnings were issued 1–2 days in advance (DWD, 2021; KNMI, 2021), these signals did not always lead to the necessary actions (Da Costa et al., 2026; Thieken et al., 2023a)."*

- Section 1 - Introduction (Line 87): the sentence introducing post-processing techniques should be improved as post-processing is generally used for other objectives, e.g. improving forecast skill, more than for estimating the uncertainty.

We thoroughly revised this paragraph based on this comment, and your comments below. For the sake of readability, we now focus only on ensemble techniques in the introduction. We introduce post-processing in Section 3.1, and added the goal of increasing forecast skill:

> *"Moreover, they use post-processing methods to correct forecast outputs based on historical forecast errors, aimed at increasing the forecast skill."*

- Section 1 - Introduction (Line 89): the sentence applies also to other variables rather than precipitation, so the authors should explicitly write that heavy precipitation is just a relevant example here: "in case some (or all) ensemble members exceed a certain threshold of heavy precipitation intensity"; for example, the same applies to hydrological ensemble forecasts or temperature.

We agree with this. We changed the sentence to clarify that the threshold can be set on other variables than rainfall:

> *"Ensemble forecasts are used to determine the probability of certain events, calculated as the share of ensemble members exceeding a pre-determined threshold (e.g. precipitation or discharge)."*

- Section 1 - Introduction (L. 92): this is an inaccurate statement as it does not reflect the current or most up-to-date situation: ensemble forecasts are already a standard in hydro-meteorology and widely used in operational systems; the statement sounds outdated and indeed it refers to a quite old reference from 2016. Since then, ensemble flood forecasting has gained significant momentum also in operational contexts, e.g., see Wu et al. (2020) and Speight et al. (2021). The authors should revise this sentence and refer to more recent studies than a 2016 one for such a statement aiming to provide a present-day perspective: "It is expected that ensemble forecasts will be widely integrated in operational forecasting chains in the near future (Pappenberger et al., 2016)." Actually, ensemble forecasts are already commonly made at most major operational weather prediction facilities worldwide (Speight et al., 2021). Most operational hydrological forecasting chains in Europe already integrate probabilistic forecasts, as also shown by the authors in this study in Northwestern Europe.

We thoroughly changed this part of the introduction, including all your suggestions:

> *"Over the last decade, ensemble techniques have been increasingly adopted in operational forecasting centres (Speight et al., 2021; Wu et al., 2020), to estimate forecast uncertainty. Ensemble forecasts are used to determine the probability of certain events, calculated as the share of ensemble members exceeding a pre-determined threshold (e.g. precipitation or discharge). Forecast users can include this probabilistic information in decision-making processes, which makes them more valuable than deterministic forecasts (Richardson, 2003)."*

- Section 3.1 - Lines 245-247: the sentence can be improved for clarity and flow, e.g.: "Flanders is the only region where flood inundation forecasts are run operationally for short-term (48-hour) lead times, whereas all other regions rely on discharge forecasts and thresholds."

Thank you for rephrasing the sentence. We changed it to the sentence you proposed.

- Figure 3: The water-level symbol summarizing the number of warning levels for fluvial floods is not very clear. The authors should consider footnoting the symbols or explaining them in the caption to help readers' comprehension. Moreover, the font size of the text labels appearing beside this symbol (e.g., “>100 year”, etc.) is too small and in these labels the word return period or the abbreviation RP (e.g. “>100 year RP”) should be added for clarity.

Thank you for looking to the figure in so much detail. We changed the figure and included all your suggestions, including a more detailed explanation in the caption.

- Section 3.5: typo in “a dedicated emergency management plans” - a single plan?

Indeed, this was a typo. We changed this to *“Some regions in The Netherlands have a dedicated emergency management plan for floods.”.*

- Section 3.6 (Lines 566-568): improve and clarify the following **sentence**: “Warnings for a specific colour can also be issued if there is a small chance (< 65% or < 25% of the area) on precipitation amounts belonging to the following colour.” - clarify and correct: smaller chance (or area) of precipitation amounts belonging to the next, more severe warning class? How much smaller? I guess there must be a minimum probability threshold. Also, the wording “chance on” should better read “chance of” in the following parenthesis: “a small chance on rainfall belonging”.

We clarified these thresholds in the revised manuscript. Indeed, there is a minimum and maximum range for probabilities on the next higher colour code (15%-65%), which we didn’t mention before. We checked it once more, but no minimum area is mentioned for the < 25 % area threshold. We rewrote these sentences to the following:

> “Warnings are generated using yellow-orange-red colour codes at the provincial level (Fig. 3). Warnings are issued if more than 65% of the ensembles exceed the rainfall threshold for a specific colour over more than 25% of the provincial area (Fig. 3). A warning for a given colour can also be issued if either of the following conditions is met: 1) the threshold for the next higher colour is exceeded by more than 65% of the ensembles, even if this occurs over less than 25% of the area; 2) the threshold for the next higher colour is exceeded by 15–65% of the ensembles over more than 25% of the area.”

- Section 3.6 - L. 583: 10-day forecasts should be defined as medium-range forecasts and not as ‘long-term’

This was indeed not the correct wording. Thank you for noting this. We changed it.

- Section 4 - Lines 693-694: It would be important to clarify what the following statement means: “Luxembourg implemented a new alerting system (LU-alert), harmonizing the warning levels of the meteorological forecasts alerts and the crisis management.”

We meant that the warning levels of the meteorological alerts and the crisis management alerts are the same. Therefore, we changed the sentence to:

> *“In November 2024, Luxembourg implemented a new alerting system (LU-alert) and streamlined the warning system, by adopting the same warning levels and colour codes for meteorological forecasts and crisis management alerts. This includes a new level for “immediate danger” with a purple colour.”*

- Section 4.3 - Lines 817-820: the paragraph title “Best practices of impact-based forecasting” could be removed or better phrased to represent the content, as in addition to best practices the plans and initiatives for future developments are reported.

We removed the paragraph title, including the other title in this section (“the importance of inundation forecasts”) to ensure consistency.

- Section 5 - Line 939: the word 'systems' might be missing after 'national-scale cell broadcasting'
Correct. We inserted the word ‘systems’.

References

Da Costa, J., Ebert, E., Hoffmann, D., Cloke, H. L., & Neumann, J. (2026). Signals without action: A value chain analysis of Luxembourg’s 2021 flood disaster. Natural Hazards and Earth System Sciences, 26(1), 343–366. https://doi.org/10.5194/nhess-26-343-2026

Haag, I., & Luce, A. (2008). The integrated water balance and water temperature model LARSIM-WT. Hydrological Processes: An International Journal, 22(7), 1046-1056.

Thieken, A. H., Bubeck, P., Heidenreich, A., von Keyserlingk, J., Dillenardt, L., & Otto, A.: Performance of the flood warning system in Germany in July 2021 - insights from affected residents. Natural Hazards and Earth System Sciences, 23(2), 973–990. https://doi.org/10.5194/nhess-23-973-2023, 2023a.

Speight LJ, Cranston MD, White CJ, Kelly L. Operational and emerging capabilities for surface water flood forecasting. WIREs Water. 2021; 8:e1517. https://doi.org/10.1002/wat2.1517

Wu W, Emerton R, Duan Q, Wood AW, Wetterhall F, Robertson DE. Ensemble flood forecasting: Current status and future opportunities. WIREs Water. 2020; 7:e1432. https://doi.org/10.1002/wat2.1432