

Dear Dr. Robert Sakic Trogrlic,

We would like to express our gratitude for the constructive and helpful comments from you and the two reviewers. They have contributed to major improvements in our manuscript. Please note that we propose a slight change the manuscript title, from “Comparing Flood Forecasting and Early Warning Systems in Transboundary River Basins” to “Comparing Flood Forecasting and Early Warning Systems in Northwestern Europe”, to emphasize the study location instead of the transboundary aspect.

Please find below the review comments (in black) and replies (in blue). The section, figure and table numbers refer to the revised manuscript version.

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

Tim Busker

Reviewer #1

This study reviews the status of Flood Forecasting and Early Warning Systems (FFEWSs) in transboundary river basins in the Northwestern Europe Countries that were hit by the July 2021 flood (Germany, Luxembourg, Belgium, and The Netherlands). Following the deadly and costly flood event of 2021, such analyses are essential for improving flood risk management and early warning systems chains, and to foster increased regional cooperation in transboundary river basins.

The study uses semi-structured expert interviews and literature review to analyze and compare FFEWS characteristics in the different countries, including forecast types, warning levels, communication protocols, emergency response plans and institutional coordination. Expert interviews from the region reveal that all systems are under a significant and rapid development after the 2021 flood event, which brought attention to some limitations of the FFEWS at the time. The main findings include the identification of key differences between countries and challenges, especially around harmonization and impact-based forecasting, which is still underused in the region, as only Flanders has operational inundation forecasts. Moreover, the authors find a lack of harmonization in protocols and inconsistencies in warning levels and communication protocols, which hinder cross-border coordination in transboundary river basins.

The paper is well written and contains many interesting details about the FFEWS structure in the four Countries studied, which are of interest for the community and the readers of NHESS.

However, it has some analytical limitations that should be addressed to enhance its clarity (in terms of organization of material and description of methods), rigor (in terms of **systematic comparisons**), and practical relevance (in terms of enhanced discussion and **solid recommendations**). In terms of clarity, the organization of the material can be improved as some key information is dispersed and difficult to find (see detailed comments below). A better synthesis should be made to be able to better compare all the key FFEWS characteristics across countries and regions (as further detailed in the comments below).

Moreover, the discussion of some developments, limitations and barriers of the current systems should be enhanced, **to better connect the analytical review of the FFEWS to the recommendations for their improvement.**

Thank you for your extensive and very helpful review of our manuscript, which we think increased the quality of the revised study significantly. It is nice to read that you perceived the paper as well-written and interesting to the NHSS community. We think that the changes as outlined below improved the clarity, rigor, and practical relevance of the paper, in line with your suggestions. Please find our replies to your feedback below.

Major comments

1. **More efforts towards a systematic classification and synthesis of all the important aspects of the FFEWS should be made:** some important pieces of information (like the real time data sharing across countries for both river discharge and reservoirs upstream, or the use of forecast post-processing techniques) are scattered in the text and not always reported for all countries; this makes it difficult to find the same level of information for all countries or get a synthetic view across regions. Given that some information is not included in the summary table (Table 1), for some Countries it is difficult to find all details and unclear whether an information is not reported because it is not applicable (e.g. no provisions for data sharing or no forecast post-processing are made) or whether the information was not available or not retrieved. The authors should summarize and list more clearly all the key descriptive information pieces that are currently scattered in the text and reported only as examples (e.g., L. 545 for data sharing). Possibly, a larger table than the current Table 1 (or additional tables) could make the comparisons of all the interesting aspects clearer, specifying where the information is not available, to facilitate the synthesis and interpretation of all findings.

Thank you for these helpful comments. In the new version of the manuscript we now start with a detailed comparison of the different FFEWSs systems (Section 3.1) and extended the comparative analysis by expanding Table 1 (see below). We re-consulted the forecasting and crisis management authorities to better understand the post-processing methods used operationally, and the meteorological and hydrological models used. We also organized an additional interview with the forecasting service in Wallonia (SPW). This provided new information which is integrated in different parts of the manuscript. Table 1 now includes 5 new characteristics, including the names of the meteorological and hydrological agencies. This also ensured that the provided information is consistent among the countries. For example, we realized that we did not mention the main hydrological platform used in The Netherlands (Delft-FEWS; now included in Section 3.5) and the postprocessing method used in the state of RLP (ProFoUnD analysis; now included in Section 3.3).

Data sharing is mostly organized per transboundary basin, and not per country. Therefore, we did not include this aspect in the table. However, we thoroughly revised the former Section 4.4 about data sharing. As it was not really identified as a major challenge in the interviews, we now moved part 4.4 to Section 3 (now Section 3.7). We also expanded this Section 3.7 by inclusion of two international commission's crucial for international collaboration on flood forecasting: the Internationalen Kommissionen zum Schutze der Mosel und der Saar (IKSMS) and The International Commission for the Protection of the Rhine (ICPR).

Table 1. Characteristics of Flood Forecasting and Early Warning Systems (FFEWS) in different European regions and countries.

	Country	Germany (RLP)	Germany (NRW)	Luxembourg	The Netherlands	Belgium (Flanders)	Belgium (Wallonia)
Meteo forecasts	Forecast agency	DWD		Meteolux	KNMI	RMI	
	Main ensemble model(s)	ICON-D2-EPS and ICON-EPS		ICON-(D2-)EPS, ECMWF-ENS, AROME-EPS	HarmonEPS and ECMWF-ENS	ECMWF-ENS (AROME-EPS under development)	
Flood forecasts	Forecasting agency	LfU	LANUK	AGE	WMCN and regional water authorities	HIC and VMM	SPW
	Main hydrological and model(s)	LARSIM	LARSIM	LARSIM	HBV and SOBEK	NAM and MIKE11	Hydromax
	Discharge (D) or flood inundation (I) forecast (lead time)	D (48h)	D (10 days)	D (24-48h)	D (5-days and 15-days)	D (10-day), I (48h)	D (lead time dependent on catchment size)
	Post-processing technique on discharge forecasts	ProFoUnD analysis (Haag et al., 2013).	Unknown	Yes, bias correction	Ensemble dressing (Verkade et al., 2017)	Non-parametric (Van Steenberghe et al., 2012)	Unknown
	Probabilistic forecasts?	Yes	Yes	Yes	<i>National level (WMCN):</i> Yes <i>Regional water authorities:</i> Not always	Yes	Yes
Online portal(s)	www.hochwasserzentralen.de <i>RLP:</i> https://www.hochwasser.rlp.de <i>NRW:</i> https://hochwasserportal.nrw (only observations; forecasts in preparation)		https://www.inondations.lu/map	https://waterinfo.rws.nl	<i>Flanders:</i> https://www.waterinfo.be <i>Wallonia:</i> https://hydrometrie.wallonie.be (only observations)		

Crisis management	Emergency management plans for floods	Yes	Yes, but not specifically for floods	Yes	Yes	Yes, but not specifically for floods	Yes, but not specifically for floods
	Primary warning communication system (start date of operation)	MoWas (October 2001) and DE-alert (February 2023)		LU-alert (October 2024)	NL-alert (November 2012)	BE-alert (June 2017)	

2. **The review of the FFEWS is presented in an overly descriptive way. More efforts in terms of comparative analysis and discussion should be made to move from a descriptive to a prescriptive analysis as the authors aim to do, providing some good recommendations at the end, which could be enhanced.** While the paper documents the current situation in detail, it often lacks critical analysis of why systems differ or what technical or institutional barriers may have shaped them. It also misses an opportunity to theorize and discuss barriers to impact-based forecasting, linking to known problems, e.g. governance fragmentation, models computational cost, lack of high-resolution data, trust, etc. **The authors should discuss further why the uptake of operational impact-based forecasts remains so challenging, and what technical or institutional barriers currently limit this.** Why is Flanders the only region where such forecasts are provided? If it is for the lack of inundation models with suitable computation time for operational applications? Is it for lack of institutional mandates? Why static flood hazard maps computed off-line and available as catalogues (for selected return periods, as done in EFAS) are not used? Is this known? Some limitations of the available flood maps are mentioned and quite clear (e.g., the large deficiencies of the static flood hazard maps during the 2021 flood and the need for more extreme scenarios, as discussed in Section 4.2), but the barriers for the implementation of the currently available flood maps (e.g. EFAS or other national static flood maps) or the production of more accurate inundation maps (near real-time) are not discussed sufficiently. More in general, more discussion is needed to ensure that the recommendations provided are well motivated by the found patterns and limitations. For example, the definition of the rainfall warning levels can be further discussed, as they are just given as absolute rainfall values without explaining how they were derived and why they differ (e.g. L. 160). Some more elements of analysis of the causes of the described characteristics could help, highlighting why the existing FFEWS have certain characteristics (e.g., how the rainfall thresholds are computed), and if the reasons behind these patterns (e.g. differences in rainfall thresholds) have been clarified or if the information is available at all, which is not always so clear so far. For example, the authors could clarify whether the definition of the rainfall thresholds uses different or same RPs in different countries, and why a specific RP or way of calculating is adopted. This was done better for the fluvial flood thresholds. Moreover, related to the warning levels for pluvial and fluvial floods, the authors could discuss the representativeness of the warning levels with respect to flood impacts, e.g. whether for the 2021 floods, there is any information on whether the areas exceeding the warning levels for both pluvial and fluvial floods matched well observed impacts.

As mentioned above, we took many measures to improve the comparative analysis of the paper. We agree that we did not clearly state the main barriers to develop impact-based forecasting practices. We therefore included these barriers in in the revised manuscript, at the start of Section 4.3:

“Our study confirms that the uptake of impact-based forecasting in operational systems is challenging. Most important reasons are the lack of relevant data (e.g., on flood vulnerability), institutional collaboration across sectors (e.g., between forecasting agencies and emergency management), and lack of computational power to run inundation models, especially for multiple ensemble members.”

We also rewrote Section 4.3 to better explain why inundation forecasts are so crucial for the transition to impact-based forecasting, and that impact-based forecasts can indeed be made by real-time hydrodynamic simulations, or by using pre-computed inundation maps:

“Flood inundation forecasts can be made using multiple methods, by translating streamflow to event-based flood hazard maps using a pre-computed catalogue of hydrodynamic simulations (Dottori et al., 2017), or by using operational hydrodynamic models (Apel et al., 2022).”

Although it is likely that real-time hydrodynamic simulations will offer a significant higher accuracy (e.g. due to a better representation of the hydrograph, both spatially and over time), we did not include this, as we did not ask this question during the interview, and existing studies (e.g. Dottori et al., 2017) are currently insufficiently describing the accuracy differences between both methods.

We also added more explanation on the rainfall warnings, and specifically how the thresholds are chosen (start of Section 3.2): *“Meteorological warnings are issued based on clearly defined rainfall intensity and duration thresholds (Fig. 3). The thresholds are chosen based on expected impacts to society and are different across countries. Moreover, as those impacts are highly dependent on the situation (e.g. day or night, location, large events), the thresholds are mostly used only as guidance in practice.”* The interviews did not reveal if existing impact-analysis have been done to define the thresholds, or whether thresholds were defined based on expert judgement, or a combination of both.

It would indeed be a very interesting analysis to assess if the warnings overlapped with the impacted areas, both from a meteorological and a hydrological perspective. Such analysis should be done very carefully by using historical archives of the forecasts and warnings in July 2021 and local impact data (similarly to Busker et al., 2025), and is an interesting idea for a follow-up study.

Thank you for your suggestion about improving the recommendations in the paper. We included a new section at the end of the paper (Section 5.2 Recommendations) in which we list the four main recommendations, nicely linked to the four main challenges found:

1) *Streamlined warning levels*: investigate whether harmonization of meteorological and hydrological warning levels and color codes between different countries and within different regions of a country, improves communication and decision-making.

2) *Warnings for catastrophic events*: Assess the added value of more extreme warning levels, such as the (dark) purple levels in Luxembourg and some German regions.

3) *Impact-based forecasting*: accelerate the development of operational impact-based forecasting systems. This in turn requires an expansion of resources and knowledge on impact-based forecasting, and a stronger collaboration across sectors, such as between forecasting agencies and emergency management.

4) *Forecast communication*: Implement a structural evaluation of warning communication chains, to ensure that warnings are consistently communicated and correctly interpreted by different organizations, and the local population, and ensure that the information is tailored to their needs.

The changes or progress made in the FFEWS after the July 2021 flood event are not clearly summarized and described, as they are scattered in various parts of the paper and difficult to find, e.g. in Section 3, L. 178-179 (“In Germany and Luxembourg, a cell broadcasting system was installed in response to and after the July 2021 flood.”). Moreover, there are hints at recent moves from deterministic to probabilistic forecasts, and a statement about it in the abstract (“All regions have invested in probabilistic flood forecasting systems”), while in the article these changes are not clearly reported, i.e. where and when such changes have been made.

Table 1 only reports that now all countries have probabilistic hydrological forecasts, but it is unclear when these have been established. It might be beneficial for sake of clarity to have an additional table or scheme, listing or summarizing all the recent developments in FFEWS, or including some information about recent changes in the current Figure 3 or Table 1. The information to highlight and summarize should include: (i) when and how the probabilistic FFEWS were developed (from deterministic to probabilistic or increase in ensemble size?), being this one of the key findings, (ii) when the online platforms were improved, (iii) the emergency response plans were updated, (iv) the communication protocols changed, e.g. national-scale phone-based alerts, etc. This information is only hinted at in different parts of the paper.

We agree that insufficient attention was given to the changes triggered by the July 2021 event, and that the information that was present, was scattered throughout the manuscript (also raised by reviewer #2). We thoroughly changed the manuscript to improve this aspect.

We added a new paragraph to summarize the post-2021 changes (start of Section 4):

“The interviews revealed that the FFEWSs in all regions are under strong and rapid development after the 2021 flood (Table 2). All regions now have emergency response plans for floods and use cell-broadcast services for alerting the population. A new flood forecasting team has been developed at LANUK in NRW (Table 2). Moreover, all regions defined clear thresholds for the different hydrological warning levels, except for Wallonia (Fig. 3). In autumn 2023, the flood forecasting center (LfU) in the RLP added a new warning level for very extreme floods (> 100-year return period) with a dark-purple color (Fig. 3 and Table 2, top). The flood forecasting center of Flanders (HIC) optimized their warning thresholds in close collaboration with

stakeholders, such as the Vlaamse Waterweg waterways authority, and many provinces updated their Emergency and Intervention Plans. In Wallonia, a flood expertise unit (CELEX) has been established since the 2021 floods, which can provide tailored flood impact advice during crises. Furthermore, ensemble forecasts are now used to produce flood forecasts and the web-portal for floods has been renewed. In Limburg, the flood disaster management plan (“Rampbestrijdingsplan Hoogwater Limburg”) now includes flood scenarios for small rivers (e.g. Geul river), including clear warning thresholds and corresponding actions. Moreover, a large programme has been launched in 2022 to improve flood safety: Waterveiligheid en Ruimte Limburg (WRL). In November 2024, Luxembourg implemented a new alerting system (LU-alert), harmonizing the warning levels of the meteorological forecasts and the crisis management. This includes a new level for “immediate danger” with a purple color. Even after those changes, the interviewees identified several issues that require attention to further improve FFEWSs (Table 2, bottom). We will further outline those challenges below.”.

Furthermore, we included a new Table (Table 2) to summarize those changes. We asked the interviewees to clarify the main changes after 2021 in case this was insufficiently clear. For example, we discovered that the HIC in Flanders optimized their warning thresholds after the 2021 flood, which was not clear before, and the new interview with SPW showed that they moved from deterministic to ensemble forecasting. As a result, we now have a consistent overview of the major changes for each region (Table 2). We also re-consulted the interviewees to retrieve the exact date of these changes. In the cases where the interviewees were able to answer this, we included this in the new paragraph and also in the different country sections (e.g. November 2024 – start of LU-alert and autumn 2023 for the new warning level at LfU, Germany).

The methodologies followed for the literature review and for the analysis of the interviews lack sufficient transparency and should be further clarified in Section 2.2 (Approach).

We expanded Section 2.2 by including more details of the literature review and the semi-structured interviews, which we outline below.

- For the literature review, the criteria for the selected articles (literature) inclusion are vague. A structured review approach (e.g., PRISMA or at least the search strings + database used + inclusion logic and filtering criteria) would improve transparency (and reproducibility). The authors should mention more clearly at least the search strings used, as only very general keywords are now reported in a vague way. Also, they should specify how many research articles were found and selected, how the country-specific reports were selected and how many were retrieved. We now include all the search strings used in the literature search, instead of only some examples. We did not list also all the combinations of different words, as we used a very large number of combinations. The same applies to exclusion or inclusion criteria. Furthermore, we now included the search engines used (Scopus and Google Scholar), which further improves transparency. Lastly, we included the amount of scientific peer-reviewed articles (16) and country-specific reports in national languages (14).

- For the semi-structured interviews, the paper briefly mentions that interviews were done with 13 experts, but lacks essential details on the sampling approach (Who was selected and why? How many were invited?). Also, how were the audio recordings of the interviews synthesized (automatically with AI tools or manually)?

Our sampling approach was based on the experience of the interviewee, so we ensured that all interviewees had senior positions. Moreover, we ensured that we interviewed at least 1 forecaster and 1 person directly involved in the crisis management in every country. We did not use AI tools to synthesize the audio recordings, to ensure AI biases will not influence our study and that we correctly summarize and interpret the interviews. We manually synthesized the recordings, within each of the five different FFEWS pillars. We added this to Section 2.2: *“Within each pillar, we **manually** synthesized the responses, identifying patterns, commonalities, and differences between the different organizations, based on the audio recordings obtained with the participant's consent.”*

Table 2. Most important changes of the Flood Forecasting and Early Warning Systems (FFEWS) triggered as response to the 2021 floods, and the key challenges remaining in different European regions and countries.

Country	Germany (RLP)	Germany (NRW)	Luxembourg	The Netherlands	Belgium (Flanders)	Belgium (Wallonia)
Major changes since the 2021 floods	DE-alert system, and addition of extra warning level for very extreme floods (dark purple)	DE-alert system, and establishment of a new flood forecasting center	LU-alert system, including a new warning level for “immediate danger” (purple)	A new multi-year programme on flood safety in Limburg (WRL) and updated flood disaster management plan	Optimized hydrological warning thresholds (see Fig. 3) and revised Emergency and Intervention Plans	Flood expertise units (CELEX), updated flood portal and move from deterministic to ensemble forecasting
Key challenges remaining after 2021	Streamlined warning levels and thresholds					
	Different warning levels, thresholds, and color codes are used between and within the different countries and organizations.					
	Warnings for catastrophic events					
	The highest thresholds are often too conservative (too low) and thus do not reflect major societal flood impacts.					
Impact-based forecasting						
While impact-based forecasting is recognized as crucial for efficient decision-making, such systems are not yet operationally implemented in most of the regions.						
Communication before and during the crisis phase						
Effective communication between different organizations (e.g., between meteorological agencies, flood forecasting centers, crisis management authorities) and to the population remains a challenge.						

Minor comments

- Table 1: the last column could report more clearly whether all those online platforms report publicly available forecasts and information; maybe here or in an additional column of the table, it would be interesting to summarize how the forecasts are presented, i.e. in which format (e.g., graph, map, text, etc.); also, the column heading 'Primary alerting system' could be clarified (in the heading or in the caption).

We changed 'Primary alerting system' to 'Primary warning communication system' to increase the clarity of this heading. We tried to add a new row for the content of the online portals. However, we in the end chose to not include it as the information is very similar among the different portals (graphs-maps and text is used everywhere), and the most important difference is that only NRW and Wallonia does not display their forecasts (which was already included).

- L. 62: a 'missed' forecast is less common in this context, usually this would be referred to as a 'missed' event

Indeed, this is what we meant here, so we changed it to "missed event"

- L. 70: "inaccuracies in spatio-temporal estimations ..." are mentioned, but also intensity biases should be recalled

We added "which can lead to errors in spatio-temporal or intensity estimates" to make clear that intensity is also important to evaluate.

- L. 73: the link between the sentence '*Numerical weather prediction models have improved greatly in recent decades ...*' and the following one, starting with 'For example, *the Integrated Forecasting System (IFS) operated by ...*', is not obvious and clear. It should be improved. Probably here it would be interesting to state when the IFS moved to 51 ensemble members, and why it represented an improvement with respect to previous IFS versions (deterministic or lower-size ensemble).

We revised the section accordingly. We now include when ECMWF started with ensemble forecasting (in 1992). We also include the three different systems (medium-range, sub-seasonal range and seasonal range) and their corresponding ensemble members (51, 51, and 101). In addition, we now elaborate on why ensemble forecast have greater value to users than deterministic forecasts: "*Forecast users can now incorporate the probability of a certain event as additional information in the decision-making process. Therefore, ensemble forecasts have higher value for forecast users than deterministic forecasts (Richardson, 2003).*"

- L. 80: after this sentence ('Over the last decade, ensemble and probabilistic ...') you can add at least one reference on recent reviews on hydrological ensemble forecasting

We added a paper containing an extensive review on operational hydrological ensemble forecasting systems (Papperberger et al., 2016).

- L. 84: “Thieken et al. (2023b) showed that *around a third of the flood-related fatalities in the ...*” here it would be relevant for the paper arguments to report how (on which basis) they showed this

This is a very good point, as the authors of the paper had access to official parliamentary documents, which were just to investigate whether the victims were warned or not. We added this to the manuscript.

- L. 110: it would be good to clarify immediately how (on which basis) are these recommendations developed

The methodology used (semi-structured interviews and literature research) was mentioned at the end of the paragraph. We agree it is better to mention the followed methodology immediately. We rewrote this section to:

“We intend to gain and share knowledge on FFEWSs in a transboundary context and to develop recommendations based on insights from semi-structured interviews and a literature review.”

- L. 119: I guess that ‘compromises’ is a typo, and it should be ‘comprises’

Indeed, this was a typo. We changed it to: *“which comprises of several countries”*

- L. 126-127: the issue of evacuation orders in July 2021 is specified only for Luxembourg; it is unclear and it would be relevant to specify whether to the authors’ knowledge there were no evacuation orders in the other Countries, as now the reader is left to guess so

It is very difficult to find out exactly when and where evacuation orders or advices were given. In hindsight, we also do not know for Luxembourg if this was the case everywhere. We did find evidence that at the most impacted locations (Ahr, Vesdre, Limburg) evacuation advices were mostly too late. We included this in the manuscript:

“Although official evacuation orders or advices were eventually given over most of the region, they were mostly issued too late, just before the peak floodwaters arrived, such as in the Ahr catchment (Thieken et al., 2023a, Rhein and Kreibich, 2025), in the Geul catchment in the south of The Netherlands (Endendijk et al., 2023a), and in the Vesdre catchment in Belgium (Dewals et al., 2021).”

- L. 141-142: the combinations of these keywords and the search strings should be reported in the text or an Appendix for transparency

We think it has limited added value to report all used search string combinations, but we included all main keywords used, and furthermore specified our search engines (see response to major comment 4).

- L. 142-143: it should be clarified how many reports and how were they retrieved and selected

We now include that we selected 14 country-specific reports in national languages, and that we selected these by recommendations of experts from the case study regions.

- L. 159: when did Germany add this extra warning level (dark purple) to represent events with immediate danger? It would be interesting to know in the context of the paper, highlighting developments after the 2021 flood event

Thank you for raising the question. Different than what stakeholders informed us before, it turns out that the purple warning of the German DWD was already in place years before the 2021 flood event, although it is not clear when exactly. The meteorological agency in Luxembourg (Meteolux) added a new purple warning level in November 2024. We also consulted with our partners in Rheinland Palatine (Landesamt für Umwelt Rheinland-Pfalz). Their dark purple level was implemented as a response to the 2021 flood, in Autumn 2023. We changed the manuscript accordingly (in this section, and also in Section 4.2 “*warnings for catastrophic events*”).

- L. 215: is the data exchange in real- or near real-time?

This is not specifically mentioned in the reports, and neither in other sources. However, gauging data is mostly shared in near real-time as small delays (some seconds to minutes) are frequent. We added this to the manuscript.

- L. 217: the acronym LAWA could be defined here and mentioned to which organisation it should refer (German Working Group on water issues of the Federal States and the Federal Government)?

We now defined the acronym in the text.

- L. 247: the sentence seems to suggest that no ‘simulation exercises (SimEx)’ of emergency preparedness exercises are carried out, is there any information on this?

Unfortunately we did not find information about this, so we removed this sentence from the manuscript.

- L. 267-268: “...uses different weather models from neighboring countries (ECMWF, DWD, MeteoFrance)”, ECMWF should not be mentioned here alongside National organisations, as it is not from a single neighbouring country, and actually Luxembourg is a Member State of ECMWF, so mentioning ECMWF here like this is confusing

That is a good point. However, we do think it is important to still mention ECMWF as they use those forecasts. We instead deleted “from neighboring countries” to avoid confusion.

- L. 159 and L. 272: this would be an interesting point to clarify: what exactly is the definition of an ‘imminent’ or ‘immediate’ danger, used to define the purple level in Countries where this is used?

This is indeed an event with “immediate danger”. Our interviewees clarified that these are events which require “immediate action”, such as with rapidly developing flash floods. We added this to the manuscript.

- L. 345: in the “European EFAS forecasts”, the acronym of EFAS could be reported and the reference to the Copernicus EMS of which EFAS is part. We wrote out the acronym, and also included the Copernicus Emergency Management Service.
- L. 554: here is the first time that upstream reservoirs are mentioned, in terms of data exchange; this point should be expanded a little bit, given its relevance in transboundary river basins

We agree that information sharing on reservoir levels is important. We included that reservoir levels are often not shared, specifically for energy-producing reservoirs: *“HIC in Flanders mentioned the need to improve the exchange of information about upstream reservoirs in Wallonia. Often, information of reservoir levels is not shared, especially for energy-producing reservoirs (for energy companies, such as ENGIE).”*.

Reviewer #2

This paper addresses an important and timely topic on flood forecasting and early warning systems (FFEWS) in transboundary river basins, using the 2021 flood event as a key reference point. The paper contains a wealth of interesting insights from both literature and key informant interviews (KIIs), and the forensic perspective on the 2021 disaster is particularly valuable.

However, the overall flow and structure of the paper could be strengthened to help the main arguments and contributions emerge more clearly. In particular, the introduction and Section 3 would benefit from more explicit framing, stronger transitions between paragraphs, and a clearer delineation between pre- and post-2021 developments. The methodology also appears somewhat light and would benefit from revision to ensure there is a more systematic approach to data collection, content analysis and communication.

We want to thank you for your time and effort to review the manuscript. It is very nice to read that you think the results are interesting and valuable. We carefully responded to the issues you raised, which are outlined below. In our opinion, this led to a strong improvement of the manuscript. The section, figure and table numbers refer to the revised manuscript version.

Major comments

1. Overall structure and flow

- The literature review provides valuable forensic insights from 2021, but the narrative currently mixes several issues in single paragraphs. More explicit structuring could help the key challenges and gaps stand out.

We carefully revised the introduction. We restructured the introduction into 3 different paragraphs: 1.1 A Flood Forecasting and Early Warning System (FFEWS), 1.2 Current challenges in FFEWS and 1.3 Research gap and aim. This required the revision and relocation of large pieces of text, including sometimes to other sections. Such movements and revisions include the part about the flood hazard zones, (which is now included in Section 4.4), the inclusion of additional information about the lack of uptake on IbF, and an expansion of the research gap.

- The introduction highlights important lessons but does not yet bring out the main research gap clearly.

Thank you for pointing this out. We did not realize that we did not clearly explain the research gap. Therefore, we now included the research gap in Section 1.3: *“Given these challenges, it is crucial to compare and assess the current state of operational FFEWSs. Only very few studies assess operational FFEWSs in detail (e.g. Kaltenberger et al., 2020; Schroeter et al., 2021), and to the best of our knowledge none of them*

assess all components of the FFEWSs chain, including communication and crisis management. Moreover, a transboundary analysis (e.g. on different warning levels in different countries) is lacking.”

- The flow from paragraph to paragraph can be strengthened with more topic sentences and transitions that guide the reader through the logic. The introduction currently mixes a range of issues in one paragraph; separating them more clearly could help highlight the specific challenges the paper aims to address.

We have rewritten large parts of the manuscript to improve the structure and flow between the paragraphs. As an example, we added the following lines to improve the connection between different parts of the manuscript:

- Start of Section 3: *“In this section, the key characteristics of FFEWS are presented based on recent literature and the expert interviews. First, we present a general overview of the different systems (Section 3.1), after which we discuss the warning levels used (Section 3.2), a detailed overview per country (Sections 3.3-3.6), and finally provide an overview of collaborations between the different regions (Section 3.7).”*
- Right before Section 4.1: *“Even after those changes, the interviewees identified several issues that require attention to further improve FFEWSs (Table 2, bottom). We will further outline those challenges below.”*

As mentioned in response to your previous comment, we also rewrote the Introduction, to improve the challenges identified and the key knowledge gaps. Moreover, we added sub-headings to all main sections (Section 1-5) of the manuscript to improve the readability.

- It remains unclear whether Section 3 is purely descriptive (based on literature) or includes empirical data from KIIs. Clarifying this distinction is essential.

Section 3 is based on a combination of both literature and information obtained in the expert interviews. To clarify this, we added the following to the start of section 3: *“In this section, the key characteristics of FFEWS are presented based on recent literature and the expert interviews. “*

2. Definition of research gap and aim

- The research gap and aim of the paper should be defined more clearly and earlier on.

We improved the definition of the research gap by including the following (Section 1.3):

“Given these challenges, it is crucial to compare and assess the current state of operational FFEWSs. Only very few studies assess operational FFEWSs in detail (e.g. Kaltenberger et al., 2020; Schroeter et al., 2021), and to the best of our knowledge none of them has assessed all components of the FFEWSs chain, including communication and crisis management. Moreover, a transboundary analysis (e.g. an international comparison of warning levels) is lacking.”

We understand that you would prefer to have the research gap mentioned earlier. However, as NHESS is not a dedicated forecasting journal, we think it is important to first explain the core principles of FFEWS, before moving into the research aim. Although we did not move the research aim to an earlier section, the new structure (with sub-headings) allows readers to easily jump towards the research aim if they understand FFEWS already.

- The research question and focus area can also be stated more explicitly, ideally near the end of the introduction.

We mentioned the research questions in the last part of the introduction. However, we did not clearly state the focus areas of our research. Therefore, we included this at the end of the introduction: *“We focus on the areas hit by the July 2021 flood, specifically The Netherlands (Limburg), the Rhineland-Palatinate (RLP) and North Rhine-Westphalia (NRW) federal states in Germany, Belgium (Flanders and Wallonia) and Luxembourg.”*

3. Clarity of arguments in early sections (L83–L95)

- From L83 onwards, the paragraph discusses communication issues but then shifts to examples where modelling outputs were inaccurate (e.g., flood zone delineation). These examples appear to relate more to forecast accuracy than to communication.

We mentioned the flood zone delineation issue to illustrate that people outside the official zones have a lower risk perception, and that therefore, warnings should be communicated differently to ensure that also these people take action. We understand that this was not clear. Therefore, we have rewritten the paragraph and restructured it to point out to the issues of 1) warning dissemination (i.e. not everyone received a warning, 2) communication issues (related to risk perception and behavior), and 3) cross-border collaboration and harmonization.

- The discussion around L90–L95 needs better alignment: the statement on flood awareness between in- and out-of-floodzone populations seems inconsistent with earlier points about fatalities outside the delineated zones, suggesting that flood extents exceeded forecasts.

With “flood zones” we referred to the legally defined flood zones in spatial planning policies, prior to the flood, and not to the forecasted flood (extent) itself. We understand that this was not clear, so we clarified it by changing “flood zone” to “the defined static flood hazard zones (as defined in spatial planning policies)”. We hope that the paragraph in its current form is clearer.

- The sentence on adaptation motivation does not connect directly with the statement on flood warning access (L95).

With the changes mentioned above, we think this is clear now.

4. Clarification of key terms and assumptions

- L165: Please elaborate on what constitutes a “clearly defined alarm level.” When is this not clearly defined? It seems this may relate to objective levels corresponding to forecast thresholds and expected impacts.

That’s a good point. We now included that with clearly defined we mean “based on specific discharge or water level thresholds”. We changed this in the new version of the manuscript.

5. Section 3: Presentation and organization

- The paragraphs describing the table and figure are difficult to follow. Consider adding more guiding sentences to help the reader navigate these visuals.

We rewrote the first paragraphs of Section 3 to a large extent, including more guiding sentences (see reply to comment 1 for some examples). We also restructured this section by dividing the first parts into two new sub-sections: 3.1 Overview of the operational FFEWS and 3.2 Warning levels in the region. As a response to reviewer #1, we also expanded Table 1. We included a detailed explanation of this table, in Section 3.1, to provide the reader with extra guidance.

6. Depth and rigor of the methodology

- The methodology section seems quite light: the sample size is small, and there is no indication of systematic coding or content analysis.

We performed a thorough literature review and combined this with expert interviews, which is a proven method to conduct assessment studies. However, it is a good point that the method description could be enhanced. This is why we expanded the methodology section (Section 2.2: Approach) by providing more information on the literature search (e.g. search engine and keywords used) and the number of articles/reports selected (16 scientific articles and 14 national reports). Of course, a much higher number of reports have been read before the selection took place. Furthermore, we interviewed 13 experts. Although this is indeed not a very large sample size, we took a number of measures to ensure that the results are valid. First, we ensured that we interviewed at least 1 forecaster and 1 person directly involved in

the crisis management in every country. This is added to the manuscript. Second, the interviews were very productive and lasted long (mostly >2 hours), which ensured that the information was interpreted correctly. Third, almost all interviewees reviewed the paper before submission to validate the results and the co-authors themselves are also experts across all study regions. Therefore, we are convinced that the gathered results reflect reality and provide a holistic view on the current state of FFEWS. We analysed the content based on a thematic analysis, conducted in every pillar of the early-warning chain. We describe this in the middle of the section.

- For research question (b), a more in-depth analysis of communication materials would strengthen the conclusions.

In Section 4.4, we used 15 different sources to show why communication challenges exist (e.g. missing communication protocols, lack of actionable information, fragmented pieces of information from multiple stakeholders). We have rewritten this section, and we think that in the current form it gives a clear view on the challenges and how those can be addressed. Furthermore, we added a clear recommendation about improving communication (Section 5.2): *“Forecast communication: Implement a structural evaluation of warning communication chains, to ensure that warnings are consistently communicated and correctly interpreted by different organizations, and the local population, and ensure that the information is tailored to their needs.”*

7. Treatment of transboundary dynamics

- The transboundary challenges could be brought out more clearly, particularly regarding data sharing and alignment of alert levels.

Because of your comment, we realized that we did not focus enough on the transboundary aspects of data sharing. Therefore, we expanded the paragraph on transboundary data sharing and now list the most important transboundary commissions which facilitate data exchange (Section 3.7). Our interviewees actually did not mention many challenges in transboundary data sharing but rather emphasized that this is already quite well organized. Therefore, we moved this section from Section 4 (“Key challenges remaining after the 2021 flood”) to Section 3 (“Key characteristics of the operational FFEWS”).

We also revised Section 4.1, mainly to emphasize that the differences in alert levels are not only present within different regions of a country (and between the meteorological and hydrological agencies), but also between the different countries.

- Consider including another figure to illustrate the communication side of these systems.

This is a very nice idea, but we think this can only be included in a proper way in a follow-up study. The manuscript is already quite long (especially with the new Table 2) and therefore we think an extra figure (including explanation) better suits a follow-up study dedicated to forecast communication.

8. Evaluation of effectiveness

- The paper provides a rich description but limited critical evaluation of the accuracy and effectiveness of the new developments in FFEWS.

In Section 4, we provide a detailed description of the challenges that remain after the upgrades triggered by the 2021 flood event. In the old manuscript, it was maybe not clear that this section addressed the remaining issues after 2021. Therefore, we renamed the section to “4. Key challenges remaining after the 2021 flood”. Moreover, some parts of the section have been expanded. We cannot really provide clear conclusions yet on the effectiveness of the new developments after 2021, as we will need to be careful with that and the only way you can really know this is after a large flood happened in the area.

- It would be useful to reflect on whether improvements have been validated and to synthesize recommendations based on that assessment.

We synthesized the issues remaining after the 2021 flood in four key challenges. We now mention these challenges clearly in Table 2. We also developed four concrete recommendations to address those challenges, one for every challenge. These recommendations are listed at the end of the manuscript, in Section 5.2:

1) *Streamlined warning levels*: investigate whether harmonization of meteorological and hydrological warning levels and color codes between different countries and within different regions of a country, improves communication and decision-making.

2) *Warnings for catastrophic events*: Assess the added value of more extreme warning levels, such as the (dark) purple levels in Luxembourg and some German regions.

3) *Impact-based forecasting*: accelerate the development of operational impact-based forecasting systems. This in turn requires an expansion of resources and knowledge on impact-based forecasting, and a stronger collaboration across sectors, such as between forecasting agencies and emergency management.

4) *Forecast communication*: Implement a structural evaluation of warning communication chains, to ensure that warnings are consistently communicated and correctly interpreted by different organizations, and the local population, and ensure that the information is tailored to their needs.

In line with our response above, it is good to mention that a validation of the improvements can only reliably be done after a flood event took place.

9. Scope and temporal framing

- The content currently mixes pre- and post-2021 developments, leading to ambiguity about the study’s temporal focus. Clarify whether the analysis primarily concerns pre-2021 systems, post-2021 changes, or both.

Thank you for this very useful comment. We totally agree that this was a weak spot of our previous manuscript version. Therefore, we took several measures to improve the definition of the study’s temporal focus. First, we renamed Section 4 to “*Key challenges remaining after the 2021 flood*” to stress that those are the challenges remaining after the improvements triggered by the 2021 flood. Secondly, we included a new table (Table 2, see below) to list the major changes of the FFEWS triggered by the 2021 flood in the different regions. We also describe all those changes in the beginning of Section 4.

Minor Comments

- **L1, L119:** Write out abbreviations such as “bn” and “mm.”

Thank you. This is done.

- **L83–L95:** Improve alignment between discussion of communication issues and modelling inaccuracies.

Done (see comments above on point 3)

- **L165:** Clarify “clearly defined alarm level.”

Here we meant “based on specific discharge or water level thresholds” (see response on point 4)

- **L492:** Clarify the source of the quoted text.

The quoted text originates from the interviews we had in Wallonia. We changed it to: “*Our interviewees in Wallonia stated that “...there should be a new discussion about the need for an extra level of warning (e.g., dark red) for really extreme situations.”*”

- **L495:** The phrase “cope for decision makers” is awkward—consider revising to “decision-making challenges” or similar.

Done

- **Throughout:** Review paragraph transitions and ensure topic sentences clearly indicate the purpose of each paragraph.

As you can see in our track changes document, we thoroughly revised the entire manuscript, including new paragraph transitions, a better structure and improved writing style.

Table 2. Most important changes of the Flood Forecasting and Early Warning Systems (FFEWS) triggered as response to the 2021 floods, and the key challenges remaining in different European regions and countries.

Country	Germany (RLP)	Germany (NRW)	Luxembourg	The Netherlands	Belgium (Flanders)	Belgium (Wallonia)
Major changes since the 2021 floods	DE-alert system, and addition of extra warning level for very extreme floods (dark purple)	DE-alert system, and establishment of a new flood forecasting center	LU-alert system, including a new warning level for “immediate danger” (purple)	A new multi-year programme on flood safety in Limburg (WRL) and updated flood disaster management plan	Optimized hydrological warning thresholds (see Fig. 3) and revised Emergency and Intervention Plans	Flood expertise units (CELEX), updated flood portal and move from deterministic to ensemble forecasting
Key challenges remaining after 2021	Streamlined warning levels and thresholds					
	Different warning levels, thresholds, and color codes are used between and within the different countries and organizations.					
	Warnings for catastrophic events					
	The highest thresholds are often too conservative (too low) and thus do not reflect major societal flood impacts.					
Impact-based forecasting						
While impact-based forecasting is recognized as crucial for efficient decision-making, such systems are not yet operationally implemented in most of the regions.						
Communication before and during the crisis phase						
Effective communication between different organizations (e.g., between meteorological agencies, flood forecasting centers, crisis management authorities) and to the population remains a challenge.						

References

- Apel, H., Vorogushyn, S., & Merz, B.: Brief communication: Impact forecasting could substantially improve the emergency management of deadly floods: case study July 2021 floods in Germany. *Natural Hazards and Earth System Sciences*, 22(9). <https://doi.org/10.5194/nhess-22-3005-2022>, 2022.
- Busker, T., van den Hurk, B., de Moel, H., & Aerts, J. C. (2025). The value of precipitation forecasts to anticipate floods. *Bulletin of the American Meteorological Society*, 106(3), E473-E491.
- Dewals, Benjamin, Erpicum, Sébastien, Pirotton, Michel, Archambeau, Pierre: Extreme floods in Belgium. The July 2021 extreme floods in the Belgian part of the Meuse basin, 2021. <https://henry.baw.de/server/api/core/bitstreams/6b2b809e-becd-465e-bd3b-d24e3a9f467a/content>
- Dottori, F., Kalas, M., Salamon, P., Bianchi, A., Alfieri, L., and Feyen, L.: An operational procedure for rapid flood risk assessment in Europe, *Nat. Hazards Earth Syst. Sci.*, 17, 1111–1126, <https://doi.org/10.5194/nhess-17-1111-2017>, 2017.
- Endendijk, T., Botzen, W., de Moel, H., Aerts, J., Duijndam, S., Slager, K., Kok, M.: Experience From the 2021 Floods in the Netherlands: Household Survey Results on Impacts and Responses. *Journal of Coastal and Riverine Flood Risk*, 2. <https://doi.org/10.59490/jcrfr.2023.0009>, 2023a.
- Haag, I., Ehret, U., Demuth, N., & Moritz, K.: ProFoUnD: Ein Werkzeug zur statistischen Analyse von Abfluss-und Wasserstandsvorhersagen und zur Ermittlung von Vorhersageunsicherheiten, 2013.
- Kaltenberger, R., Schaffhauser, A., & Staudinger, M.: “What the weather will do” – results of a survey on impact-oriented and impact-based warnings in European NMHSs. *Advances in Science and Research*, 17, 29–38. <https://doi.org/10.5194/asr-17-29-2020>, 2020.
- Pappenberger, F., Pagano, T. C., Brown, J. D., Alfieri, L., Lavers, D. A., Berthet, L., Bressand, F., Cloke, H. L., Cranston, M., Danhelka, J., Demargne, J., Demuth, N., de Saint-Aubin, C., Feikema, P. M., Fresch, M. A., Garçon, R., Gelfan, A., He, Y., Hu, Y.-Z., Janet, B., Jurdy, N., Javelle, P., Kuchment, L., Laborda, Y., Langsholt, E., Le Lay, M., Li, Z. J., Mannessiez, F., Marchandise, A., Marty, R., Meißner, D., Manful, D., Organde, D., Pourret, V., Rademacher, S., Ramos, M.-H., Reinbold, D., Tibaldi, S., Silvano, P., Salamon, P., Shin, D., Sorbet, C., Sprokkereef, E., Thiemig, V., Tuteja, N. K., van Andel, S. J., Verkade, J. S., Vehviläinen, B., Vogelbacher, A., Wetterhall, F., Zappa, M., Van der Zwan, R. E., and Thielen-del Pozo, J.: Hydrological Ensemble Prediction Systems Around the Globe, in: *Handbook of Hydrometeorological Ensemble Forecasting*, Springer, Berlin, Heidelberg, https://doi.org/10.1007/978-3-642-40457-3_47-1, 2016.
- Van Steenbergen, N., Ronsyn, J., Willems, P.: A non-parametric data-based approach for probabilistic flood forecasting in support of uncertainty communication. *Environmental Modelling & Software*, 33, 92-105, 2012.

Verkade, J. S., & Werner, M. G. F.: Estimating the benefits of single value and probability forecasting for flood warning. *Hydrology and Earth System Sciences*, 15(12), 3751-3765, 2011.

Rhein, B. and Kreibich, H.: Causes of the exceptionally high number of fatalities in the Ahr valley, Germany, during the 2021 flood, *Nat. Hazards Earth Syst. Sci.*, 25, 581–589, <https://doi.org/10.5194/nhess-25-581-2025>, 2025.

Richardson, D. S.: Predictability and economic value. Seminar on Predictability of Weather and Climate, 9-13 September 2002, 321–332. <https://www.ecmwf.int/node/11922>, 2003.

Schroeter, S., Richter, H., Arthur, C., Wilke, D., Dunford, M., Wehner, M., and Ebert, E.: Forecasting the impacts of severe weather, *The Aust. J. Emergency Manage.*, 36, 76–83, <https://doi.org/10.3316/informit.766653620579430>, 2021.

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.