RC2: 'Comment on egusphere-2023-907', Anonymous Referee #2, 08 Aug 2023

General comments:

This article presents a comparison between different QPF-forced hydrologic predictions of flash flooding events, featuring a recently developed ensemble technique at Meteo France. Selected events are used to assess the ability of these predictions to detect the occurrence of streamflow values exceeding defined frequency-based thresholds. QPE-forced hydrologic simulations are used as reference data. Metrics based on contingency table data such as the Critical Success Index are used to assess skill. While I don't see any new scientific insights into QPF-forced hydrologic forecasts, this is a good contribution to the literature of tools with operational implementation. The article's methods are based on robust work previously published by others. The manuscript is written well and in a concise manner. My only concern is that the authors do not provide enough details about the featured technique that is central to the study. My recommendation is that following some minor revisions, the paper could be accepted.

⇨ We thank Referee #2 for the careful reading of our manuscript and the relevant comments and suggestions. We provide point to point answers below, including details about the way we plan to adapt the manuscript.

• My main comment is that the PIAF-EPS methodology is not described with enough details. The workflow schematic in Figure 2 does not provide any information on how the perturbations are computed, which seems to be an important aspect of the technique. Authors should include an example of these perturbations, so readers can get a sense of what they look like.

⇨ We propose to add the following figure, line 137: *"An example of the perturbations is given in figure 3."*
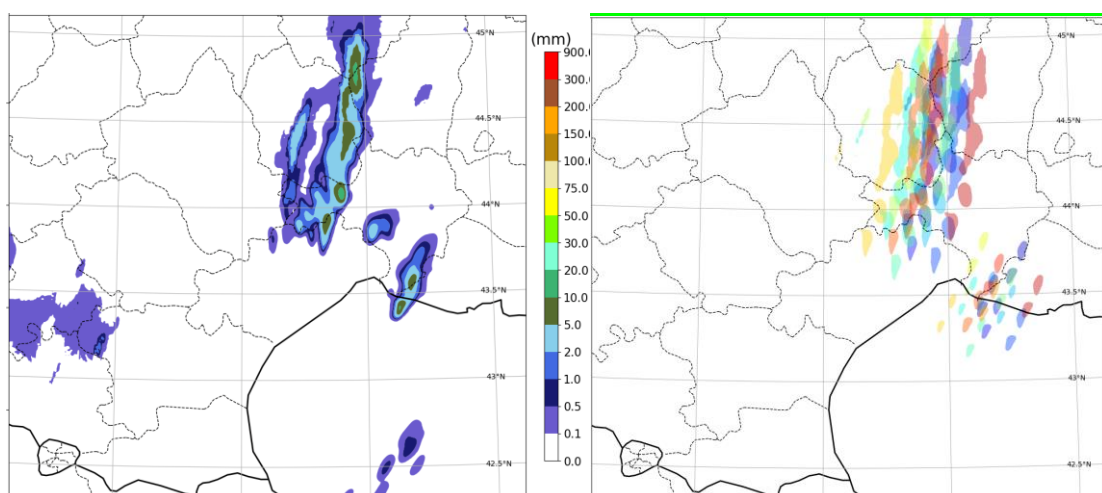
Figure 3: example of PIAF-EPS ensemble forecast perturbations. Left : deterministic PIAF forecast of 15-minute rainfall accumulation (forecast start: 19 Sept 2020 at 06utc, forecast range: 2 hours). This is used as member zero of the ensemble. Right: same field in members 1 to 16, the shading represents rainfall areas above 5mm, with one colour for each member.

- Many acronyms not spelled out the first time they appear in the text. At least some of them are spelled out later in the document, but they should be spelled out as soon as they are used for the first time, so the reader is not left wondering about it.

⇨ Thank you for noticing this. We will carefully check that acronyms are spelled at first occurrence.

Specific comments:

Line 61: "…for flash flood nowcasting purposes" Is it appropriate to say "flash flood nowcasting"? I have only seen nowcasting being used to describe QPE extrapolation.

⇨ We agree to rephrase as "*for flash flood forecasting purposes*"

Line 81: "SMASH" (L80) Is "PANTHERE" an acronym? If so, please spell it out.

⇨ SMASH acronym is detailed in section 2.4, and PANTHERE is indeed an acronym. As a consequence, the sentences will be modified as follows: "*The simulated/forecast hydrographs and the reference discharges are obtained using a fully distributed rainfall runoff model, detailed in section 2.4. In the operational version of Vigicrues Flash, this hydrological model is forced with the PANTHERE (Projet Aramis Nouvelles Technologies en Hydrométéorologie Extension et Renouvellement) rainfall QPEs, derived from a network of about 30 radars over mainland France and its vicinity (Tabary et al., 2013).*"

Line 94: What is "…a ten-minute observation cutoff"? Do you mean only the first ten minutes worth of observations are assimilated? How many observations (how many data points) are actually assimilated? Also, specify what data are assimilated (radar, satellite, rain gauge?).

⇨ We will insert the following explanation on line 94: "*...with a ten-minute observation cutoff (i.e. the initial state of each forecast is prepared using observations collected up to 10 minutes after its validity time)*." Also, additional information will be inserted at line 95: "*Each 3D-Var analysis updates the model state by multivariately blending tens of thousands of observations from various meteorological networks (including radar winds and reflectivities, satellite radiances, GPS data, in situ surface and aircraft reports, etc). More information about the AROME-NWC 3D-Var can be found in Auger et al (2015).*"

Line 96 – 97: Spelling out this acronym (PIAF) should occur earlier in the document, as soon as it is first used. Same with all other acronyms.

⇨ Line 60 (first occurrence of "PIAF") will be modified as follows: "*The objective of this paper is to assess the potential of a new seamless short-range ensemble QPF product, called PIAF-EPS ("PIAF" meaning Prévision Immédiate Agrégée Fusionnée, and "EPS"*

meaning Ensemble Prediction System) and recently developed by Meteo-France, for flash flood nowcasting purposes."

Line 92 – 112: Use of "Lead time". Consider replacing the term "Lead time" with something like forecast length, or simply referring to a particular forecast by its length. For example, the 3h forecast, to refer to a forecast that goes out 3 hours into the future. The term "Lead time" implies skill associated to a particular forecast length, and not a configurable parameter.

⇨ Agreed, we will replace "lead time" by "forecast range" which is universally used in the meteorological community ("forecast length" is rather used for the total forecast duration, which is a different thing). In this study, the forecast length is 3 hours and the forecast range can take values between 15 minutes and 3 hours.

Line 119: With "equiprobable", do you mean perturbations are "drawn" from a uniform probability distribution?

⇨ No, the meaning is "equiprobable: having the same degree of logical or mathematical probability" (www.merriam-webster.com). The distributions are clarified a few lines after: 2D Gaussian sample for spatial perturbations, and clipped AR(1) autoregressive process for the amplitude perturbations. Sentence will be clarified as follows: "*using a priori equiprobable perturbations of the precipitation field, as explained below*:"

Line 121: What do you mean with "subrandom"?

⇨ We will correct as "pseudorandom"

Line 126: A better term to replace "lead time" here would be forecast length.

⇨ We will replace by "... as a function of forecast range"

Line 172 – 173: How costly? How often is the system changing?

⇨ We will rephrase *as "(it would be labour intensive to process older cases, because of technical constraints in the archiving system, and they would be less and less relevant to current operational forecasting systems because the AROME and PIAF systems are frequently upgraded, typically once a year)"*

Line 176: "…the importance of hydrological **reaction** response"?

⇨ We propose to rephrase this sentence as follows to be more clear and explicit: "*.. and the intensity and geographical extent of the hydrological responses simulated by the SMASH model*"

Table 1: "Duration" does not seem appropriate. Use "Date"? Also, I see the order of the events in here is by date, but labels "A-H" are all over the place. Not a big deal, but this is very odd order to follow here and Figure 5. It feels like the labels' purpose was to make it

easier for the events to be organized/classified, but the way they are presented in this table seems to defeat said purpose?

⇨ *The table will be re-organized, and "Duration" will be replaced by "Date".*

Line 211 – 212: The word "assimilated" is misleading. Do you recursively use reference streamflow to improve model states and/or parameters? If not, then I strongly recommend using a different term here.

⇨ *"assimilated" will be replaced by "assigned".*

Line 255: Why was the 60th percentile chosen for the comparison? I could not see anything in the previous texts that would give indication that a particular percentile was to be used.

⇨ The sentence starting in line 254 will be modified as follows: *"Moreover, the 60% percentile of the PIAF-EPS forecast, which has the highest CSI score among the other percentiles, shows even better results than the PIAF forecast CSI=0.27), by reducing notably the area affected by false alarms."*

Line 360 – 361: More than confirming, a robust study on a large enough sample dataset should inform how truly valuable and applicable is the ensemble-based technique, particularly in real-time.

⇨ We agree, such additional study on a large and continuous sample would bring additional information, not only confirmation. The sentence will be modified as follows: "*The results presented here should nevertheless be complemented with more robust statistical evaluations over longer periods of time and on a larger number of high precipitation events, bringing a more generic overview of the quality of the forecast ensembles.*"