Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC3
https://doi.org/10.5194/nhess-2022-182-RC3, 2022


# Comment on nhess-2022-182

Anonymous Referee #3

---

Referee comment on "A methodological framework for the evaluation of short-range flash-flood hydrometeorological forecasts at the event scale" by Maryse Charpentier-Noyer et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2022-182-RC3, 2022

---

## Overview

The manuscript shares its focus between the verification of accuracy of ensemble precipitation forecasts and different ways to convey (and analyse) the information provided in terms of discharge forecast by a meteo-hydrological forecasting chain. On the one hand, the theme of forecasting severe rainfall events is largely discussed in the introduction, but a in-depth analysis on the verification of output by NWP models (and related ensembles) is neglected. On the other hand, it is declared that a new framework for the evaluation of meteo-hydrological model coupling is proposed, but a proper review of past studies about this issue is not provided in the introduction and the proposed analysis recalls (and put together) different approaches commonly used in the operational practice of worldwide flood forecasting centers. In addition, many parts of the proposed evaluation framework appears as unsuitable for real-time applications. The overall feeling about the present manuscript is that it describes a very detailed post-event analysis, where the parts of novelty and originality do not clearly stand out. A clear choice about the main goal of the study should be taken and then properly developed. In my opinion, the strong point of the paper should be the availability of three meteorological ensemble products (even though it is not clear if a performance comparison for these ensemble is a novelty or past studies have already investigated the subject). The performance evaluations in terms of Quantitative Precipitation Forecasts (QPFs) should be based on a larger dataset and taking into consideration the concept of “fuzzy verification”. The analysis of outcomes provided by a meteo-hydrological model chain driven by the available ensemble QPFs is an added value for the study.

## General comments

1) The main declared aim of the manuscript is to presents a methodological framework for the event-based evaluation of ensemble forecasts for floods, with respect to the needs of civil protection authorities. But, the proposed analysis is quite complex (several aspects

and score to consider), maybe not suitable to the real-time operational practice of flood forecasting centers. The current contents of Section 4 sound more like a post-event analysis. In addition, the use of verification metrics like rank diagrams and ROC curves to analyse a single event has poor significance (Figs. 7 and 8). These metrics are commonly used over large datasets, in order to highlight statistical characteristics of the forecast product. The computation over a single event could be of some interest if compared to "historical" performances based on a long archive (for instance, for real time applications, the spread skill relationship in Figs. 11-16 does not add significant information with respect to the issue of warnings and outcomes shown in the remaining panels).

The statistical analysis in terms of discharge forecast should consider the whole period covered by QPFs (not just a flood event).

The coupling with an hydrological model represents a complementary tool for the verification of QPFs (since catchments can be seen as macro-raingauges with variable interception areas), given that the intermittence of the rainfall signal is dampened by the non-linearity in rainfall-runoff processes. In particular, the dynamics of the overall soil filling and depletion mechanisms and the flood routing play a fundamental role in determining results, as well as the role of the morphology of the basin that determines the time-space scale below which the variability of the rainfall field is dumped. The spatial integrating effect of a watershed filters out some of the spatial and temporal variability that complicate the point-by-point verifications that are more commonly used (Benoit et al., 2000).

2) The proposed evaluation of rainfall forecast is aimed to take into account spatial and temporal variability. The proposed analysis recalls in a some way the concept of the so-called "double-penalty effect" (i.e., the fuzzy verification introduced by Ebert, 2008 and Roberts and Lean, 2008, and discussed by Schwartz and Sobash, 2017). But the subject is treated neglecting specific past literature about this issue. Introduction and Section 4.1 should be revised accordingly.

Why has just one year of ensemble forecasts been used, given that products are available from 2018?

3) AROME-EPS and AROME-NWC with time lagging are merged to build an ensemble. Which are the reasons to merge the two products? Why is AROME-NWC with time lagging just used to build an ensemble?

4) The reasons for using two different hydrological models for different aims should be discussed

5) Contents of Section 4.3 should be reformulated taking into account the response times of the considered catchments. Outcomes depends concurrently by the accuracy of rainfall forecast for the event study as well as by the characteristic of the basin.

**Specific comments**

Line 7: “peak flood” in place of “flood rising limb”, given that the statistical analysis is focused on the maximum value of the discharge forecast

Lines 15-17: this statement is questionable due to the limited dataset; results do not support “to draw robust conclusions”. A reformulation needs.

Lines 69-72: this content (i.e., point i) ) recalls what has mainly been done in this manuscript

Lines 72-73: this content (i.e., point ii) ) is questionable by the light of general comment 2)

Lines 89-91: this content should be revised taking into account the general comment 2).

Lines 119-121: this subject should be deeper investigated in the introduction.

Lines 130-132: this subject should be deeper investigated taking into consideration the concept of fuzzy verification

Lines 133- 153: the proposed analysis and metrics fits well for a post-event analysis but they are not suitable for real time operational practices, with respect to the point of view of end-users.

Lines 145-146: this statement is questionable, given that an evaluation of performance based on the last hours is not indicative about the performance of hourly QPF in the following future time-steps

Lines 147-153: the use of rank diagrams to analyse a single event appears as no fully proper

Lines 157-162: these contents can be simply summarized by stating that the forecast is verified within a time window useful for the aims of end-users (warning issues)

Line 164: how is the 10-yr return period computed for the ungauged basins?

Lines 215-222: the use of observations which were not available in real time to calibrate the hydrological models limits the operational use of the proposed forecasting chain. As well as the peak discharges estimated at ungauged locations during a post-flood field campaign makes impossible to replicate the proposed framework for real time applications.

Line 233: if AROME-EPS is updated every 6 hours, it is not clear how figs 5, 6, 7, 11-16, B1-B6 show continuous hourly forecast with 1 to 6 hour lead times for each hourly time step.

Line 245: “an ideal distance for the present case study” fits better than “an ideal distance”.

Lines 248-253: is this comparison a novelty with respect to past literature? Why the 1-h lead time is not considered to build Fig.3? Additional rainfall accumulations larger than 5 mm/h should be considered to complete Fig.3.

Line 257: the description of the use of each model within the present study should be here introduced.

Line 270: specify the periods of the calibration and validation processes

Line 271: define the acronym KGE

Line 281: specify the time step at which this model runs

Line 291: specify in the text the period of the temporal evolution; what does J+1 mean?

Line 292: the 1-h lead-time has poor significance for the aims of end-users (i.e., warning issues). The 3-h lead time is more significant.

Caption Fig.4: define the acronym Hymex (or avoid to use it in the caption)

Line 304: it is not clear to what "rising limb" is referred

Lines 305-309: for certain selected outlets, hyetographs for 6-h rainfall amount (for a fixed or moving average time window) should be also useful to evaluate the impact of rainfall forecast on the hydrological forecast, due to the integrating effect of the spatial-temporal variability of rainfall by the rainfall-runoff processes

Lines 312-315: this statement is questionable, given that, in real-time, it is not known which areas will not contribute, even if a nowcasting forecast is available. The different scales involved between model predictions and raingauge measures, coupled with the high variability of the physical events and of the model errors, complicate the use of precipitation observations for atmospheric model validation, particularly in complex terrain endowed with a limited density of instruments. This areal variability enables to diagnose different problems associated with the atmospheric simulations, such as the quality of the larger scales simulated or the reliability of the description of small scale processes. The dependence between basins and sub-basins can be very useful to understand the possible problems of spatial shifting in the modelled atmosphere (Benoit et al., 2000; Jasper and Kaufmann, 2003).

Line 319: the 1-h lead-time has poor significance for the aims of warning issues (observed rainfall plays the major role in the modelled basin response for this lead time). The 6-h lead time is more significant.

Line 332: the comment for line 319 is valid also here

Caption Fig.6: "amount" in place of "rates"

Lines 345-350: these considerations should be done on the discharge ensemble (not on the ensemble QPFs), due to the non-linearity in rainfall-runoff processes

Line 360: which is the need to run the model at 15-min time resolution?

Lines 364-365: how is the 10-yr return period computed for all the sub-basins (I guess that many of them are ungauged basins)?

Line 370: which hydrological runs were used to built Fig.8?

Caption Fig.8: specify what represent the points on each line

Line 372: which is the starting time of RFO? How long is the RFO run driven by observed rainfall before the rainfall is set to zero?

Lines 376-377: specify in the text the number of missed detections (as done for false alarms at line 379)

Line 381: "contrasted effects" is not clear to what refers to.

Lines 383-386: these considerations could be misleading (the non-linearity in rainfall-runoff processes plays a major role; it is not an effect of what percentile to consider)

Caption Fig.8: are river gauge level available every 15 minutes? How can hits be computed everywhere with a 15-min time step?

Line 388: “rainfall forecast products” in place of “ensemble rainfall forecast products” (given the general validity of the sentence)

Lines 390-392: the impact of RFO depends on the concentration time (i.e., the response time of the watershed to the rainfall) of the considered basins. Related false alarms decrease with the lead-time increasing (except for systematic errors in the hydrological simulation).

Fig. 10: in the labels, the word “Ensemble” is not clear to what refer to

Line 394: false alarms and misses should be also evaluated as function of different anticipation times

Lines 403-406: the question is doubtful, given that the concentration time strongly influences outcomes and the corresponding evaluation.

Lines 409-414: which is the sense of the analysis in terms of PC? Is PC computed in the same way used for scores shown in Fig.9?

Lines 416-419: this sentence highlights the limit of the present manuscript, given that the proposed framework cannot be applied in real-time

Lines 419-421: this statement has no sense (with respect the aims of flood warning). The accuracy of the rainfall forecast influences the quality of the hydrological forecast, but the use of RFO cannot be considered an alternative solution.

Line 425: quantify the size of the catchments related to outlets 1 and 2

Line 426: have outlets 1 and 2 weak reaction to rainfall in general or just for this event?.

Line 429: quantify the size of the catchments related to outlets 3 and 4

Line 432: quantify the size of the catchments related to outlets 5 and 6

Line 441: briefly recall the definition of the spread/skill score and specify if it is referred to rainfall or discharge forecast

Line 443: the choice of the lead-time should be appropriate to the concentration time of the investigated catchment to analyze outcomes. Otherwise, the outcomes seems to depend on the lead time of rainfall forecast

Line 445: wrong label for the outlet number in Figs of appendix B (outlet 4 for all the graphs)

Lines 462-465: maybe, the outcome is affected by a spatial scale of the shift which is not optimal for the investigated catchment

Line 471: which is the concentration time for these outlets?

Line 481: typing error

Line 495: this outcome is likely influenced by the concentration time of investigated outlets

Fig.14: in all the graphs, move the legend panel in order to do not cover lines of results

Lines 525-527: the reasons for this outcome are the same cited at line 514 (influence of the concentration time of investigated outlets). Reformulate the sentence.

Fig.16: in the graphs b) and c) move the legend panel in order to do not cover lines of results

Lines 548-554: a map displaying concentration times of investigated outlets satisfy the need. The threshold anticipation maps in Fig.9 describe just a case study related to the specific case study and forecast products. It cannot be used in general terms for flood warning purposes.

Lines 557-561: the meaning of “anticipation time” may be misunderstood. It derives by a combined effect of accuracy of currents forecast and response time of the investigated outlet.

Lines 575-579: this analysis is significant when performed over a long dataset

Lines 584-595: the sense of these considerations is related to the role of QPFs in general, not specifically to ensembles.

Lines 596-601: the gain is due to NWC. Which is the added value to use NWC+EPS rather than just NWC?

Lines 619-621: maybe, the extension of 20 km is not the optimal dimension for the investigated case study. An investigation about this issue is worth to be performed.

Line 639: have authors considered to apply spatial perturbations just to NWC members?

Lines 727-728: these contents may be misleading. The outcome depends specifically on the accuracy of the ensemble for the case study. It is not an information that can be estimated a priori by means of a statistical analysis and generally related to the lead time. It is strictly related to the investigated event and selected run of the ensemble. These contents can be referred just to a post-event analysis (and cannot be inferred for real-time operational practices).

**References used in the review comment**

Benoit R, Pellerin P, Kouwen N, Ritchie H, Donaldson N, Joe P, Soulis E (2000) Toward the use of coupled atmospheric and hydrologic models at regional scale. Mon Wea Rev 128: 1681–1706

Ebert, 2008: Fuzzy verification of high resolution gridded forecasts: A review and proposed framework. Meteor. Appl., 15, 51–64, doi:10.1002/met.25

Jasper K, Kaufmann P (2003) Coupled runoff simulations as validation tools for atmospheric models at the regional scale. Q J R Meteorol Soc 129: 673–693

Roberts, N. M., and H. W. Lean, 2008: Scale-selective verification of rainfall accumulations from high-resolution forecasts of convective events. Mon. Wea. Rev., 136, 78–97, doi:10.1175/ 2007MWR2123.1

Schwartz, C. S. and Sobash, R. A.: Generating probabilistic forecasts from convection-allowing ensembles using neighborhood approaches: A review and recommendations, Mon. Weather Rev., 145, 3397–3418, https://doi.org/10.1175/MWR-D-16-0400.1, 2017.