NHESS-2023-192

Authors' Responses to Reviewer 1 (RC1, anonymous)

Date: 2 September 2024

Title: Investigation of an extreme rainfall event during 8–12 December 2018 over central Viet Nam – Part 2: An evaluation of predictability using a time-lagged cloud-resolving ensemble system

Authors: C. C. Wang et al.

Firstly, **we thank the reviewer for spending valuable time reviewing the paper and giving us constructive comments that helps to improve the clarity of the paper.**

COMMENTS

In general, I think this is a good follow-up study to evaluate the usefulness of CReSS ensembles in forecasting an extreme rainfall event over Viet Nam. However, I believe significant editorial improvements (texts and figures) are necessary to enhance the manuscript's readability. For example, varies data are described in the data section (Section 2.1), but many of them are not used in the subsequent sections (e.g. TIGGE, WRF). Some of the colour bars and their labels are quite small. Overall, I think the current manuscript needs major revision.

Specific comments:

Abstract: The authors should state the full form first before using abbreviations e.g. CReSS, QPF, SSS, and ESA.

Reply: Thank you for your comment. We have made corrections and give the full form in the first use of these abbreviations. The corrections are shown on page 1, lines: 17, 20 and 21, in the marked-up manuscript version.

Line 36: Perhaps it would add some clarity to "The observational data …" by adding GPM in the text.

Reply: Thanks for your suggestion. Here, we are referring to the rain-gauge observations, so we mention this explicitly to add some clarity, along the lines as

suggested. The GMP data are also used in this study, so they are mentioned later in the text.

Lines 40-42: Missing reference(s)?

Reply: We agree with you, we have added the reference as suggested. The reference was added on page 2, line 44, in the marked-up manuscript version.

Lines 97-98: Missing reference(s)?

Reply: Thank you for your concern. The references were cited on line 101 in the previous draft, and we now move some of them here for better fluency and clarity, along the lines as suggested. The references were added on page 5, line 103, in the marked-up manuscript version.

Lines 113-119: I am not sure whether these details are relevant to the current study.

Reply: Thank you for your comments. To avoid too much detail, we have deleted the specific numbers in the favorable factors in the revision, as suggested. The changes are shown on page 6, lines 119-124, in the marked-up manuscript version.

Lines 123-139: Given the nature of this study is to demonstrate the CReSS ensemble can produce good forecast of extreme precipitation events. Perhaps it is necessary to have a brief description of other models, which could perform simulation at a cloud-resolving resolution, e.g. MM5 by Son and Tan (2009), WRF by Toan et al. (2018) and Nhu et al. (2017). Information such as resolution and relevant parameterisation schemes (if any) would be relevant in this case. If the resolution and relevant parametrisation schemes of those models would be the same as the current study, the authors might want to emphasis this point.

Reply: Thanks for your suggestion. We will make changes and add the brief description in the revision, as suggested. The changes are shown on pages 6 -7, lines 130-132, lines 136-139, and lines 146-147, in the marked-up manuscript version.

Lines 139-147: The authors might want to highlight the fact that these global NWP models do not have "cloud-resolving resolution".

Reply: Yes, you are right and we will make the highlight as suggested. Besides, we would also like to point out that the current global weather prediction models cannot capture the precipitation at the correct quantity in Vietnam. The added

information is shown on page 7, lines 158-160, in the marked-up manuscript version.

Line 173: Not sure why UKMO is mentioned but their forecast outputs were not used. Also "ECMWF of the European Union" is not accurate as the UK is a member state of ECMWF but not in the European Union.

Reply: Thank you for pointing this out. We have removed that information. We used the UKMO data in the first version of the manuscript. However, we had removed it from the submitted version of manuscript. But we were missed to remove the relevant information at line 173. The changes are shown on page 10, lines 218-219, in the marked-up manuscript version.

Lines 175-176: The authors should state the usage of this dataset at the beginning of the paragraph first rather than in the middle of the paragraph.

Reply: Thanks for your suggestion. In the revision, we have mentioned that this dataset is used in the study in the first sentence of the paragraph, as suggested. The changes are shown on page 9, lines 212-213, in the marked-up manuscript version.

Section 2.1.2: It might be worthwhile to highlight the fact that NCEP GFS is a deterministic forecast.

Reply: Thank you for your comments. As suggested, this clarification has been made in the revision and now the sentence reads as "… from the analyses and deterministic forecast runs…" the change is shown on page 12. Lines 272-273, in the marked-up manuscript version.

Section 2.1.3: The authors should highlight the fact that these are the in-situ observation data as GPM data is also a type of "observation" data.

Reply: Thank you for your suggestion. We will make change in the revision, as suggested. The changes are shown on page 8, lines 181-182, in the marked-up manuscript version.

Section 2.1: The authors might want to restructure this section so that the data, which serves as similar purpose would be grouped together. A possible structure could be: "Section 2.1.1" In-situ observation data and "Section 2.1.2" GPM are used to model validation (ground truth); "Section 2.1.3" TIGGE and "Section

2.1.4" WRF are used to demonstrate the added values of CReSS ensemble; "Section 2.1.5" NCEP GFS is used to drive CReSS. This structure would fit nicely into "Section 2.2 Model description …"

Reply: Thank you for your suggestions. We will make change in the revision, as suggested. The changes are shown on pages 8 -12, in the marked-up manuscript version.

Figure 1, Section 2.1: ERA5 was used in Figure 1 but was not mentioned in Section 2.1.

Reply: Thank you for your comments. After carefully checking, we found that Fig. 1f where ERA5 data are used, has been adapted directly from Fig. 14c of Wang and Nguyen 2023, and we also stated this information in the caption of Figure 1. Therefore, we did not mention it in section 2.1. if our explanation is not legal, we are happy to make the change, in the marked-up manuscript version.

Lines 228-230: "The first members ran at 12:00 UTC on 3 December 2018, and the last member ran at 12:00 UTC on 10 December 2018…" à "The first member was initialised at 12:00 UTC on 3 December 2018, and the last member was initialised at 12:00 UTC on 10 December 2018. A new member was initialised every 6-hr within the period 1200 UTC 3 Dec 2018-1200 UTC 10 Dec 2018."

Reply: Thank you for your suggestions. We will make change in the revision, as suggested. The changes are shown on page 12, lines 281-285, in the marked-up manuscript version.

Section 2.1.2, Lines 232-234, and Table 1: I am a bit confused. Did the authors use NCEP FNL (Lines 232-234, Table 1; also, from the first part of this two-part study [Wang and Nguyen 2022]) or NCEP GFS (as stated in Section 2.1.2) to drive CReSS?

Reply: Thank you for your comments. We corrected the information. It is NCEP GFS data as stated in Section 2.1.2. The correction is shown on page 12, lines 288-289, in the marked-up manuscript version.

Line 278: "If small spread …" à "For example, small spread indicates…"

Reply: Corrected. The correction is shown on page 16, line 336, in the marked-up manuscript version.

Figure 3, Lines 312-313, Figure 5: I believe the OBS, which the authors are referring to, is not from GPM nor in-situ observations but from ERA5 as it has more similarity to Figure 1f than Figure 1e. The authors also mentioned stations but stations are not indicated in Figure 3. Perhaps, the authors meant to show another figure?

Reply: We apologize for the confusion. The data shown in Fig. 3 OBS is a combination of 24-h observed rainfall from the rain-gauge network and the surface wind analysis from ERA5, and the same is shown in Fig. 1f. We have clarified the information in the caption of both Figs. 1 and 3.

Figure 3: Perhaps I missed it, but it is not clear to me how good members (green) and bad members (red) are determined. For example, why does "00:00 UTC 9" is a good member whereas "18:00 UTC 8", "06:00 UTC 9", and "12:00 UTC 9" are not classified as good members but the spatial structures of these members are very similar to "00:00 UTC 9".

Reply: We apologize for the confusion. In this study, good and bad members are determined by both visual comparison and the SSS skill score. Specifically, we compared the model results with observation, and good members were picked if the quantitative rainfall and the spatial distribution of the rain band were closer to observation. For the skill score, we applied the similarity skill score to all members and picked good or bad members based on their score. Higher score means better members and lower score means worse member.
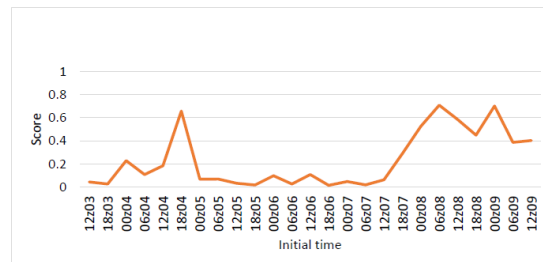
Figure 1. Similarity skill score of 25 time-lagged members for the 24-h accumulated rainfall of the Dec 10, 2018

Lines 318, 324: "… executed …" à "… initialised …"

Reply: Corrected. The correction is shown on page 17, lines 377-378, and 382, in the marked-up manuscript version.

Line 341: Is it GFS or FNL?

Reply: Here, we are referring to the GFS forecast data, so it is the GFS instead of the FNL data, which is the analysis. Specifically, it is GFS version ds084.1. We hope that this is clear.

Lines 353-367; Figure 5: (1) Perhaps it might be clearer if the authors would rename the subgroups by using the period of forecast initialisations, e.g. "First 4 members" à "12 UTC 3- 06 UTC 4 Dec". (2) Similar grouping could be done by defining period using a "moving window period", e.g. "Subgroup 1: 12 UTC 3- 06 UTC 4 Dec", "Subgroup 2: 18 UTC 3- 12 UTC 4 Dec" etc.. The "moving window" approach could help the authors to better locate the optimal initialisation periods. (3) If the "moving window" approach is used, perhaps some kind of measure of the deviation of the 24h rainfall field between OBS and subgroups would be useful in quantifying the optimal initialisation periods.

Reply: Thank you for your comments and suggestions. We will add more information to our analysis to make it clearer. The added information is shown on page 20, lines 417, 421, and 424, in the marked-up manuscript version.

Line 365: The rainfall of second 4 members is the highest "in" these… (A word is missing)

Reply: Thank you for your careful review. We have corrected it, as suggested. The correction is shown on page 20, line 424, in the marked-up manuscript version.

Line 366: … due to "a single good forecast initialised at 1800 UTC on 4 Dec".

Reply: Thank you for your careful review. We have corrected it, as suggested. The correction is shown on page 20, line 425, in the marked-up manuscript version.

Lines 384-388: The definition of subgroups seems to be arbitrary. It is not clear to me why these subgroups are chosen for in-depth analysis. Furthermore, would this not be expected that the forecasts initialised closer to the target period would have a better forecast skill as certain features that are highly related to the extreme precipitation event are included in the initial conditions? In this sense, an interesting question would be: Why does the forecast initialised on 1800 UTC 4 Dec can (partially) capture the spatial extent of the extreme precipitation event of interest and such information was lost for the next 72 hours?

Reply: Our purpose in this paragraph is to examine how forecast quality evolved as the lead time shortened. Therefore, along with the evaluation on time-lagged results using batches of fixed number of successive runs (every 4 members) as presented in our analysis, this study also grouped the members using different ensemble sizes based on their behavior in order to better assess the temporal evolution of forecast uncertainty and event predictability as the lead time shortened

Figures 10, 11, 12: The authors should add some labels (u, v, qv) on the plot. Also missing labels on Figure 10a and Figure 10c

Reply: Thank you for your comments and suggestions. We will add this information to our figures to make it clearer. The added information is shown in Figures 11-13, in the marked-up manuscript version.

Figure 10: Is it relating to surface wind, which typically is defined as 10-m wind, or 100-m wind (as stated in Lines 473-474).

Reply: Thank you for your comments. After carefully checking, we found that the mentioned data is exactly at the surface level. We corrected the information and this correction is shown on page 31, line 599, in the marked-up manuscript version.

Section 3.2: The authors should indicate which graph/panels the readers should be looking for. E.g. Lines 476-479: It should be referring to Figure 10a-c. etc.

Reply: Thank you for your comments and suggestions. Here, we are referring to wind components in Figures. 10a-f. We have added more information to our analysis to make it clearer. The added information is shown on pages 31-32, lines 605, 609, and 615, in the marked-up manuscript version.

Lines 502-509: I am confused about the Figures and the texts. What levels are actually showing and the authors are referring to. I guess the authors are comparing Figures 11 and 10? If this is the case, Line 504 should include something like "… 100 meters (Figure 10).

Reply: Thank you for your comments and suggestions. We also apologize for the confusion. Here, we are referring to wind components and water vapor mixing ratio at the surface. Therefore, the Figure referred to is Figure 11, as you guessed. we have added more information to our analysis to make it clearer. The added information is shown on page 33, lines 641, in the marked-up manuscript version.

Authors' Responses to Reviewer 2 (RC2, anonymous)

COMMENTS

Substantial Comments (Comments are not listed in order of importance):

Comment 01. I find the evaluation of predictability for this event incomplete. Grouping ensemble members together is a very superficial evaluation of predictability that does not go into the depth needed to examine the actual cause of the lack of predictability. The ensemble sensitivity analysis is a logical next step, but it is simply presented in the manuscript. The results are not interpreted or physically linked back to the event. They are simply presented.

Reply: Thank you for your comments. We are reinstructing our paper to make our current study more logical and clearer. Specifically, we moved Fig. 4 to section 3.2, adding more new results, as well as put more emphasis on the analysis to (1) better clarify the predictability of D18 event, and to (2) present the new findings in our study more effectively and more clearly. The change is shown on pages 19-20, lines 393-409, and in section 3.2, in the marked-up manuscript version.

a. The result of predictability increases as lead time decreases is not a new result to the body of literature on meteorological prediction.

Reply: You are right. However, our results once again reconfirm that predictability decreases as lead time increases in general, especially for a region with complex terrain and weather systems like Vietnam. Furthermore, in our study, it is shown that the lead time is not the only factor to influence the predictability (as one may suspect). Specifically, there is a good forecast at a longer lead time, made by the member ran at 1800 UTC 4 Dec. This indicates the potential for heavy rainfall to occur with more time for preparation. We will put more emphasis on this member in the revision, and conclude that it is still possible to have good forecasts at a lead time up to 5 days. The deeper analysis is shown in section 3.2. pages 29-31, in the marked-up manuscript version.

b. An analysis of the differences, physically, between each ensemble run that might be the cause of the lack of predictability should be undertaken. This should be more than just low-level RH and surface winds, as it is well known that large scale features are important to controlling these factors (see the results from Part 1). This will then lend context to the ESA and identify how these sensitivities feedback into the prediction.

Reply: Thank you for comments. Based on the results from Part 1, the low-level wind convergence led to moisture convergence and these conditions resulted in the D18 event. Furthermore, the southward movement of the low-level wind convergence also dictated the movement of heavy rainband during the event. Therefore, we would like to put more focus on these two aspects in the revision.

To clarify your concern, we are plotting more results from NCEP FNL analysis data. We would like to compare the evolution of synoptic-scale patterns (features) in the sole good member at a longer lead time (at 1800 UTC 4 Dec) and bad member with that in NCEP FNL analyses to better link the performance of CReSS runs back to the physics, and also better point out a new interesting result in this study. that is still possible to have good forecasts at lead time up to 5 days. The new result is Figure 10, in the marked-up manuscript version.

Besides, we also plot the sensitivity at t-36 to better identify the timing with higher predictability. Because we found a new interesting result (new finding) in section 3.2 that the synoptic pattern already developed into (or toward) what would cause the rainfall later at timing more than 24h earlier. We believe that these could explain the predictability of D18 event. The plots for the sensitivity at t-36 are added in Figures 11-13, in the marked-up manuscript version.

Comment 02. No hypotheses are presented in this work. This leads to the manuscript being unorganized, and the results unclear in the context of the broader literature. Having model simulations are not alone publishable. It is thus important to outline scientific based hypotheses in which the experiments in the manuscript are designed to evaluate, which will then make it clearer how the work adds to the body of literature.

Reply: As we replied to your comment 01 that we are reorganizing the manuscript to better explain the predictability of D18 event as well as the performance of the CReSS model. We also add more results in the revision, as replied to comment 01, to have deeper look at interesting results (new findings), including (1) it is still possible to have good forecasts at longer lead time (up to 5 days) for record-break events like D18 event. (2) the development of synoptic patterns into (or toward) what could lead to an extreme rainfall later at timing earlier in the past. By these changes, we believe that our present study will have valuable contributions to the body of literature. The change is observed in the marked-up manuscript version.

Comment 03. Ensemble spread is not purely error or a representation of accuracy. The goal of a well calibrated ensemble is to represent the forecast probability density

function. Thus, if there is high uncertainty, we want the ensemble to have a large amount of spread. If there is small uncertainty in the system, we want the ensemble to have little spread. The usage of spread as an error metric needs to be done within this context.

a. It is also not clear to me where the spread analysis is undertaken within the paper.

Reply: We apologize for the confusion. In the time-lagged approach, only one run is executed at each initial time, so for each time it is not possible to derive the probability information. However, when successive runs are grouped together, one can see that the spread (e.g., standard deviation) still evolves with time just like a multi-member ensemble, as shown in Fig. 8. As the lead time shortened, the high SD region from the last several members became more focused along the coast, indicating high rainfall amounts (also in Fig. 9) there but with uncertainty in exact locations.

b. Some discussion and framing of the work here from a context of intrinsic versus practical predictability is needed. Additionally, the scale dependence of predictability. I suggest Melhauser and Zhang (2012), Nielsen and Schumacher (2016), Weyn and Durran (2018), and citations within as starting points. There is also some useful suggestions from an ensemble analysis within these papers.

Reply: Thank you for your comments and your information. We will add more discussion (as mentioned above) and cite these references in the revision.

Comment 04. The results presented in this paper would have made an interesting section in Part 1 paper but because of these issues outline above it is not in the current state publishable on their own.

Reply: Thank you for the comment. As it is, the Part I paper has already been published, and we certainly would like to published the high-resolution time-lagged QPF result on the D18 event as Part II, if at all possible. We will devote efforts to clarify the predictability, factors affecting the predictability, and the usefulness of the time-lagged ensemble in this event. We hope that our reply (including those stated above) and future revision will shed light on the above issues, be satisfactory, and make contributions to the existing literature on heavy-rainfall QPFs.

Additional Comments (Comments are not listed in order of importance):

1. Lines 21: typo "predicts"

Reply: Thank you for your comment. We corrected it. The correction is shown on page 1, line 22, in the marked-up manuscript version.

2. Lines 53-55: The phrasing of this sentence is awkward. I recommend removing "until now" and adding something like "to improve predictability" to the end of the sentence.

Reply: Thank you for your suggestion. We updated it, as suggested. The change is shown on page 3, line 57, in the marked-up manuscript version.

3. Lines 85-87: Citations are needed to support this statement.

Reply: Thank you for your suggestion. We added the information, as suggested. The added information is shown on page 5, line 91, in the marked-up manuscript version.

4. Lines 182-184: What version of the GFS?

Reply: we used the GFS version ds084.1. We also added the information in the revision for clarification. The added information is shown on page 12, line 272, in the marked-up manuscript version.

5. Section 2.1.3: Is there any citation or information about the observational error associated with the gauges in this network?

Reply: Thank you for your question. The automated rain-gauge network in Vietnam is operated, maintained, and managed by Vietnam National Centre for Hydro-Meteorological Network (NCN). Therefore, the network meets the World Meteorological Organization standard. The observed dataset in this study is provided by NCN through the Mid-Central Regional Hydro Meteorological Center.

6. Section 2.1.4: A citation needs to be added for IMERG. Additionally, please let us know what version you used.

Reply: Thank you for your comment. We used the IMERG Final Run V07 data production. We also clarified this and added a citation, as suggested. The change is shown on pages 8-9, in the marked-up manuscript version.

7. Line 247: Table 2 does not appear anywhere in the manuscript.

Reply: Thank you for your comments. We had removed that information. We created the table 2 in the first version of the manuscript. However, we had removed it from the submitted version of manuscript. But we were missed to remove the relevant information at line 247. The change is shown on page 14, lines 304-306, in the marked-up manuscript version.

8. Lines 485-488, Lines 508-509: It is also possible the convective inflow to the storms is not at the surface but is elevated. Again, a much more detailed examination into the variables that control the ingredients for extreme rainfall is needed in the ensemble runs and ESA.

Reply: You are right. It is necessary to add more detailed examinations. Therefore, we will put more emphasis on the analysis along with adding new results to figure out the quantitative contribution of selected variables to the D18 event. We will let you see it in the revision. The change is shown in section 3.2, in the marked-up manuscript version.

9. Section 3.2: What does "per SD" mean?

Reply: We apologize for the confusion. It means per standard deviation. We will add more information along with quantitative values in the revision for clarification. The change is shown on pages 32-35, lines 611, 671-672, 679, in the marked-up manuscript version.