NHESS-2020-397
Authors' Responses to Reviewer 2 (RC2, anonymous)
Date: 2 August 2021

Title: Evaluation of Mei-yu Heavy-Rainfall Quantitative Precipitation Forecasts in Taiwan by A Cloud-Resolving Model for Three Seasons of 2012–2014

Authors: C.-C. Wang, P.-Y. Chuang, C.-S. Chang, K. Tsuboki, S.-Y. Huang, and G.-C. Leu

1. Overall comments:

This paper evaluates the performance of a convection-permitting model (the Cloud-Resolving Storm Simulator; CReSS) in simulating heavy precipitation over Taiwan during three mei-yu seasons in 2012, 2013 and 2014. The simulations are validated against rain gauges, radar data and NCEP analyses. A well-chosen classification criteria for the events is suggested, dividing the data into segments depending on whether at least 10% of the gauges showed heavy, moderate, some or no rain. For the validation, the authors employ several verification metrics based on contingency tables, namely Threat Score (TS), Probability of Detection (POD), False Alarm Ratio (FAR) and Bias Score (BS).

Although the methods and findings of the manuscript are not totally ground-breaking or unknown to the scientific community, the study presents a compelling analysis of Quantitative Precipitation Forecasts (QPFs) validation. Provided the evaluated model works at a convection permitting resolution and the investigation region is prone to suffering heavy precipitation events with high impact, the manuscript can be relevant to scientists working on the topic as well as stake-holders dealing with the impacts of such events. Therefore I believe there is interest in its publishing although it requires major revisions.

Reply:

The positive view and constructive comments from this reviewer (Reviewer 2) are deeply appreciated, and the paper has been revised according to the comments from all reviewers. In the revised manuscript (color-coded version), the changes made in response to Reviewer 1, Reviewer 2, Prof. G. T.-J. Chen (community comment), and by ourselves (mostly minor changes in English) are marked in red, blue, green, and orange, respectively. A point-by-point response to each of the comments from this reviewer are given below following their order. In each point, how and where the revision is made in the text is also specified.

2. General comments:

1) The conclusions in Section 5 (also in the Abstract) should be further elaborated. There is no need in these sections to write again the numbers of TS, FAR, etc. (e.g. lines 554 to 555 or 563 to 564). By doing so the main conclusions of the paper are hidden e.g. that QPFs are clearly improved by the use of convection permitting in heavy precipitation situations. Please, rewrite the conclusions and the Abstract to clearly point out the findings of the paper.

Reply: Thank you for this suggestion. In Section 6, it is revised to convey the findings of this study more clearly to the readers, as suggested (L469-491). The section is now partitioned into three paragraphs, each stating the findings linked to the specific aim and purpose. In the revision, the actual score values are also reduced to just twice, one for the day-1 overall TS (all events) and the other for the day-1 TS for A+ group, to illustrate the dependency property and the scores for extreme events, also along the lines as suggested (L472-473; L478). Similarly, the abstract is also revised (L15-16) as suggested.

2) It is not demonstrated that the CReSS model does a better job in orographic precipitation (phase locked) situations compared to transient systems as its stated in the conclusions (Lines 573 to 574). I agree that presumably, predictability is larger in those situations, due to 1) the fact that stationary systems have larger intrinsic predictability and 2) the better representation of the model orography. However the paper does not demonstrate this aspect. The concept “predictability” is lightly used and no quantification is provided (see for instance Hochman et al., 2021 where intrinsic predictability is quantified using dimension and persistence metrics for the systems). Instead only one case is shown (09-10 Jun) and two more cases are mentioned (L480) but no results are provided. From the case shown (09-10) the larger predictability of orographic precipitation is assumed by the fact that the location of the precipitating front (convective line) is not well represented but the TS scores are high (TS=0.4). This is not sufficient proof. It is advised that, given the type of information and analysis provided, the paper focuses on the “accuracy” of the simulation avoiding the analysis on “predictability”.

Reference:

Hochman, A., Scher, S., Quinting, J., Pinto, J. G., and Messori, G.: A new view of heat wave dynamics and predictability over the eastern Mediterranean, Earth Syst. Dynam., 12, 133–149, https://doi.org/10.5194/esd-12-133-2021, 2021.

Reply: Thank you for this suggestion and we agree with it. As pointed out by this reviewer, the concept "predictability" is lightly used and no quantification is provided in this paper, it has been revised to focus on the "accuracy" of the simulation and avoid the analysis on "predictability" as suggested (L486-487). The reference of Hochman et al. (2021) is cited (L577-579), and some of the views of this reviewer are also incorporated into the text in the revision (L407-408). It is also better clarified in the revision that in our example case (Section 4), since the CReSS model produces most of the rainfall over 300 mm in the mountain regions (Fig. 6), the majority of the hits in this event must also occur in such regions at and above this threshold, along the lines as suggested (L327-329). Therefore, the evidence indicates that the hits occur mostly in the mountain regions, and this conclusion is not deduced just through the fact that the frontal location in our example is not well predicted.

3. Specific comments:

1) Title and Abstract (L18): It is highly possible that the reader is not familiar with what the Mei-Yu season is and when it occurs. Please include this information.

Reply: In the revision, it is added that the mei-yu season is May-June in the abstract, as suggested (L14).

2) L20-L21, L26-27, L71-73, L79-80: A perfect forecast would show a TS of 1. Why are TS values close to 0.1 a good result then? To support this statement either provide the information about the skill for that score or provide the TS values of the "past results and 5-km models". Since at this point of the paper the TS has not yet been defined please also include the information that a perfect forecast has a TS=1.

Reply: While the perfect value is 1, what constitutes a good TS value is mainly based on past results in the literature and experience of the operational sector (Dr. Leu, our last author, is the director of the Meteorological Forecast Center at the CWB). Some of these previous studies (e.g., Chien et al., 2002, 2006; Chien and Jou, 2004; Yang et al., 2004) are cited in the text to provide a proper context (L47-50). Also, it is revised to provide the range of TS ($0 \leq TS \leq 1$) at its first appearance as suggested (L173-174).

3) L88. The "dependency property" regarding the link between large events and improvement of the QPFs has not yet been explicitly explained. If I understood correctly these are defined in the papers W15, W16. A brief explanation of what

this is, is required here.

Reply: The dependency property is explained briefly at the beginning of this paragraph, as suggested (L78-85).

4) L89: The subobjective "further evaluate the model QPFs for larger and extreme events" should be part of the purpose 1).

Reply: Moved to purpose 1) in the revision, as suggested (L87).

5) L100-101 and L113: What do you mean by "CReSS needs no nesting". Dynamical downscales always need initial and boundary conditions from other, coarser, model".

Reply: In the revision, this sentence is reworded to "… a single domain without nesting" for better clarity, as suggested (L103).

6) L104-109: More information is needed about the parametrizations used. You need to explicitly mention the shallow convection parameterization scheme and the turbulence scheme. Also referring to Table 1 and W15 and W16.

Reply: As suggested, more information is added and references to Table 1 and W15 are also made in the revision (L106-124).

7) Table 1: How is the turbulence closure treated? Some models use a TKE 1D parametrization, others use a 3D, etc. What is the case in your simulations?

Reply: It is clarified that a 1.5-order closure is used in the PBL scheme with TKE prediction in the text and in Table 1, as suggested (L110-111; L114).

8) L228: Why is TS > 0.15 the threshold for predictive skill? Please explain, also if this information comes from previous literature provide the corresponding references.

Reply: In the revision, several past studies that use a similar TS value to indicate "some predictive skill" (e.g., Chien et al., 2002, 2006, Chien and Jou, 2004; Yang et al., 2004) are cited here, as suggested (L47-50).

9) L297-303: The statement “In model forecasts, when errors grow and the evolution deviates, the chance to become less rainy (not as favourable) is higher than more rainy, more so in forecasts made earlier at longer ranges.” needs demonstration, either from literature or results.

Reply: We agree with this reviewer on this point. Thus, in the revision, this part of text is revised to “…This indicates that for larger events, the error growth with lead time in the model tends to become less rainy, as reflected in the decrease in BS…” to say only what is shown in the figure, as suggested (L259-261).

10) Figure 5: Could the authors elaborate on why are the scores lower for the events during June 2013 (either Day 1 or 2)? This aspect should also be included in the manuscript.

Reply: In the revision, several earlier studies are cited to provide likely reasons here, as suggested (L67-68).

11) Figure 7: In some panels, it seems as though the scores (TS or BS) are better on the third day of the forecast (day 3, blue line) than for the previous 2 days. Could you please explain this behaviour?

Reply: This shows essentially the dependency property, where the rainfall amount (event magnitude) apparently acts as a stronger influencing factor to the skill scores than other factors such as the forecast range. In the revision, this point is better clarified as suggested (L345-346).

12) Figure 12: The understanding of the Figure, its caption and explanation provided in the text is incomprehensible. Please rewrite. Why is the number of appearances of the different weather types, proof of the better model performance? Besides, large precipitation totals are usually linked to large-scale systems rather than localized convection, this is already known to the scientific community.

Reply: In the revision, the caption of Fig. 12 has been rewritten for better clarity as suggested (L463-467). The purpose of Fig. 12 is to link the event size (which has a positive relationship to QPF skill) to synoptic factors, as reflected by the average number of items met on the checklist used to facilitate heavy-rainfall forecasting at the CWB. Part of the relevant description is also revised to improve the readability as suggested (L453-457).

13) Summary and Concluding Remarks: Please recap the aim of the paper, main steps carried out and briefly the relevance of the study before enumerating the conclusions.

Reply: In the revision, the aim and main steps of the study are restated at the beginning of Section 6, as suggested (L469-471).

14) L556: Please, again, include what do you refer to by "compared to previous results".

Reply: In the revision, it is clarified that the "previous results" mean those reviewed in Section 1 (including Fig. 2), and several of those papers are also referenced here (L473-475), along the lines as suggested.

15) L564: When you mention "larger groups" are you referring to events with large coverage? Please reword.

Reply: Revised to "…the QPFs for larger events …" for better clarity, as suggested (L476-480).

16) L573: The statement "and such QPFs with high hit rates are clearly very useful for hazard mitigation." is not documented by your investigation. This was not shown in the paper. Delete.

Reply: Deleted as suggested (L485).

4. Writing comments:

1) L17: What does in real-time mean? Please consider deleting.

Reply: Deleted as suggested (L13).

2) L33: Should read: "Quantitative Precipitation Forecasting (QPF)…"

Reply: Changed as suggested (L32).

3) L37: Should read: "… mainly during two periods …"

Reply: Revised as suggested (L35).

4) L63-64: Delete "While the scores at the CWB will be compared with our results later," and "obviously much".

Reply: Deleted as suggested (L68-69).

5) L68-69: Model resolutions of 2.2 km are already being used in operational centres, for example de German Weather Service not only in research.

Reply: We agree with the reviewer. Here, the text is revised to "… more comparable to research, …", along the lines as suggested (L72).

6) L78: Change "more rain" by "the larger the rain".

Reply: Changed as suggested (L79).

7) L84: Wang (2015) has already been defined as W15. Please correct.

Reply: Revised similar to suggested (L84).

8) L85-86: Delete "For mei-yu rainfall in Taiwan, we are certainly keen to find out how this CRM performs, especially for the extreme events."

Reply: Deleted as suggested (L85).

9) L90-91: Delete "To answer these questions above are our objectives".

Reply: This sentence is revised according to the comment from Reviewer 1 (L86-89).

10) L113-114: Delete "which are also run four times a day, each out to 72 h (now 78 h)."

Reply: This sentence is revised according to the comment from Reviewer 1 (L119-120).

11) L116: Change: "highly dictated" by "forced".

Reply: Changed as suggested (L122).

12) Table 1: Provide the grid spacing of the topography in km as well.

Reply: Added as suggested (L114).

13) L135-136: Include this information about the relevance of the study in the abstract.

Reply: Included as suggested (L15-16).

14) L189.: Include the information that BS=1 implies no biases and that BS>(<)1 implies overestimation (underestimation) of the events.

Reply: Included as suggested (L175-176).

15) L218-236: Why is the explanation of Fig. 4 before the results concerning Figure 3?

Reply: Figure 3 has already been referenced both in the previous paragraph and in the earlier part of this paragraph (L192-201).

16) L276: Rephrase "…under-prediction for low…"

Reply: Revised as suggested (L245).

17) L295: The word "serious is not appropriate in this context"

Reply: The word "serious" is changed to "evident" along the lines as suggested (L257).

18) L447: Reword: "… June are compared…"

Reply: Revised to "The forecasts … June are compared…," as suggested (L380).

19) L496: Rephrase: "… sizes, as shown in Fig. 10…" and "… as an example…" by illustrated.

Reply: Revised to "… as illustrated in Fig. 10 … for the mei-yu regime" along the lines as suggested (L423-424).

20) L580: The sentence “it is also recommended that such events should be examined with caution and proper classification. ” does not bring any information and its colloquial.

Reply: The sentence is revised to “…, and can be helpful to hazard preparation and mitigation, along the lines as suggested (L491).