NHESS-2020-397
Authors' Responses to Reviewer 1 (RC1, anonymous)
Date: 4 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:

In this study, the authors evaluated the performance of the quantitative precipitation forecasts (QPFs) by a high-resolution CRM during the mei-yu seasons of Taiwan in 2012-2014, using categorical statistics. The results showed that the QPF skill is better for larger precipitation events, and improved compared to previous results. In addition, case analysis indicates that the strength of the high-resolution CRM lies in an improved ability to capture smaller scale processes for the phase-locked rainfall systems. These findings verify that the high-resolution CRM has good potential application in actual QPF during the mei-yu seasons of Taiwan. However, some major issues need to be clarified.

Reply:

The positive view and constructive comments from this reviewer (Reviewer 1) 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) It needs to give more explanation for the novelty of this study. As mentioned in the introduction, the purpose of this study is to clarify the dependency property in categorical scores of QPF and whether the skill of the high-resolution CRM is better than those in previous studies, although the studied object is changed from typhoons to mei-yu systems. This purpose has been basically fulfilled by W15 and W16. Therefore, they should not be considered as the novelty of this study, unless the study can prove the CReSS is sensitive to different weather

systems. However, from the conclusions, the higher-resolution CReSS primarily improved the forecast skill of the phased-locked topographic rainfall, as it better resolves the terrain and related small scale processes, which means the improvement of QPF caused by this CRM is not attributed to a better capture of the evolution of mei-yu front.

Reply: In the revision, more explanation is provided for the novelty of this study as suggested. In the introduction section, we better clarified that "the main purpose of this study is three-fold: 1) to assess the skill of the 2.5-km CReSS in predicting mei-yu rainfall at a higher resolution than before, especially for heavy to extreme rainfall events, 2) to clarify whether the dependency property in categorical scores also exists in the mei-yu regime in Taiwan? and 3) if the QPFs by CReSS prove to be improved, why or where its strength lies?" (L85-89). As demonstrated in this paper, the heavy-rainfall QPFs in the mei-yu season can be improved by using high-resolution models, and the underlying reason is also examined. We believe that these are good merits of the paper worthy of publication. While also exist in the typhoon regime, the dependency property in mei-yu regime in Taiwan has not been shown until this paper. In other places in the text, similar changes are also made to provide better context for the novelty of the study (L658-660), as suggested. We also agree with the reviewer's interpretation on why heavy rainfall QPFs are improved using a CRM, and these views are incorporated into the text in the revision (L412-414; L485), along the lines as suggested.

2) Regard the QPF skill of the CReSS on different categories of rainfall events, this study shows a better QPF skill for larger rainfall events. However, this phenomenon may also happen for other high-resolution models, as a higher resolution permits the model to capture more small scale processes to improve convection development, and thus, more rainfall production. To a certain extent, this can be indicated by Figure 3 which shows that the QPF skill of the "All" category (the black lines) has a smaller success ratio (about false alarm) than those of large rainfall categories (A and A plus; the orange and red lines) for high rainfall thresholds (such as larger than 100 mm). It means that the high-resolution CReSS not only produces larger rainfall for large rainfall events, which leads to a higher TS scores, but also produces larger rainfall for small rainfall events, which leads to a smaller success ratio. Thus, this study needs to clarify more about the advantage of the CReSS model, apart from the resolution.

Reply: We agree with the reviewer on this point. In the revision, the review's opinion in the SR (in Fig. 4) is incorporated into the text along the lines as suggested (L239-241). Also, the reviewer's interpretation on why heavy rainfall QPFs are improved using a CRM, which we agree, are incorporated into the text in the revision (L274-275), along the lines as suggested.

3) As discussed in section 4, the QPF error is also attributed to the forecast error on the evolution of mei-yu front. This error may lead to a worse QPF, as the location and timing of large rainfall can be completely incorrect. What is the cause of this error? Is it related to the boundary condition or the domain processes simulated by the CReSS? The answer of this issue can clarify the novelty of this study, as it is about the QPF associated with the mei-yu front.

Reply: In Wang et al. (2016b), the position error of the mei-yu front is presumably linked to the IC/BCs, even with a higher model resolution. This is clarified in the revision, along the lines as suggested (L396-397; L399; L401-402).

4) The comparative analysis or verification is based on a key indicator, the TS score. However, how large the value of TS score could be defined as skillful or a good skill? The study mentioned that when TS is larger than 0.15 it can be indicated "some predictive skill" (in line 228). Is there any objective definition or reference from operational prediction to support that?

Reply: As the TS is typically used to indicate predictive skill in a relative sense, we could not find a fixed value to define it as being skillful in the literature. We think, for example, that it would not be fair to say that TS = 0.20 is skillful but TS = 0.19 is not. However, based on experience of the operational sector (Dr. Leu, our last author, is the director of the Meteorological Forecast Center at the CWB) and some previous studies (e.g., Chien et al., 2002, 2006, as cited in text), a value like 0.15-0.2, which is above zero to a considerable degree, can be used to indicate some predictive skill as stated in the text. Throughout the text, it is revised to use TS more in a relative term, along the lines as suggested (L210-211).

3. Specific comments:

1) There are many places in the article that are not clearly expressed or improper use of vocabulary, which require major revisions. For examples (not exhausted), the sentences in lines 27-28 ("weaker events"->"smaller rainfall events"?), 35-40 ("where"->"when"?), 68 ("to hit"-> "that hit"?), 77 ("or event magnitude"->"or rainfall magnitude"?), 89 ("whether …" ?), 110-114 ("which are also run …"->

"which are applied for the model run …"?), 117-118 ("doubled the resolution"->"increase the resolution"?), 125-130 ("used include …"-> "used for QPF verification include …"?).

Reply: In the revision, all the above instances are corrected as suggested or modified to improve their clarity similar to suggested (L24; L36; L71; L79; L86-89; L119-120; L123; L126-128). Other modifications are also made throughout the text (in orange).

2) The manuscript uses too many abbreviations, which makes the readers hard to get the meaning of the sentences conveniently. Please delete the abbreviations which appear not frequently in the manuscript.

Reply: In the revision, the abbreviations not frequently used (such as NWP, W16, SST, VMI, … etc.) are deleted to improve the readability, as suggested.

3) Figure 3: Please explain more why after the rainfall events have been categorized into different rainfall magnitude events (A-D), different rainfall thresholds are still needed for each magnitude event.

Reply: In this study, wide range of rainfall thresholds (per 24 h) are chosen to fit rainfall events at different magnitudes. In the revision, this point is better clarified (L163; L184; L206-207), along the lines as suggested.

4) Figure 5: Why not put the CReSS results along with these model results for comparison? Are these models at a resolution of 5 km? If so, the comparison in the TS scores between the CReSS and these models is discounted, as their resolution are different.

Reply: We agree with the reviewer that the comparison in the TSs between the 2.5-km CReSS and the 5-km models is not very fair, as their resolution are different. In the revision, we therefore have moved Fig. 5 from Section 3.2 to the introduction section (and become Fig. 2) to be part of the review in research background (L91-96). This way, a direct comparison is avoided and we also stress the difference in model resolution when needed (L57-67), along the lines as suggested.

5) Lines 556-557: Did these previous results come from forecasts of an equal resolution (2.5 km)?

Reply: In the revision, it is better clarified that these results are from models at lower resolutions (L473), as suggested, and therefore the heavy-rainfall QPFs during the mei-yu season in Taiwan can be improved by using a higher resolution model like the 2.5-km CReSS. This is one of the main points of the study.