Reply to the comments of Referee # 1

Referee #2 (Anonymous)

General comments: This paper aimed to show that the performance of hydrological forecasts can be improved and the lead times of hydrological forecasts can be extended by coupling a distributed hydrological with the WRF model. Through a series of numerical experiments, the paper showed the WRF QPFs have an over-bias in precipitation forecasts. It proposed a post-processing procedure that aims to reduce the biases. It also showed that model calibration using post-processed WRF QPF can lead to better hydrological forecasting performance than using historical observed hydrological data. Finally, the paper showed that forecasts for shorter lead times have higher skills than forecasts with longer lead times. I think coupling NWP model with a process based hydrological model is a viable way to improve the skill of hydrological forecasts and to increase forecast lead times.

Reply: Thank the reviewer for his/her comments, revisions have be down based on the reviewer's comments, following are responses to the reviewer's comments one by one.

However, I found some major problems with the manuscript. First of all, how the WRF model is setup is not described at all. This is important for readers to understand how the WRF model is run and how to interpret the WRF forecasts.

Reply: The whole section 3 has been rewrote, and this issue is addressed in the revised article.

Second, the forecast results from the coupled models are based on a very limited sample of events. This makes it very hard to believe that the conclusions can be generalized. For example, the correction factors computed by the post-processing method presented in the paper is based on the three simulated events and the results can be arbitrary due to its dependence on the specific selection of the events. Therefore, I don't think the method can be generalized to other events or in real-time forecasting setting. There is a similar problem using data with a limited sample size to conduct model calibration. Model parameters must be determined based on a large sample of data for the calibration results to be robust. If parameters are optimized based on a single event, you would obtain different optimized parameters every time a different event is used.

Reply: As the study watershed is large, so it is difficult to acquire lots of flood event samples. Conclusions from this study is based on the results achieved in this article, to have a more generalized conclusion the authors agree that more studies may be needed.

For model parameter optimization, previous studies have shown for a distributed hydrological model, one flood event can be used to optimize a reasonable parameters, so the parameter set achieved in this study can be used for other flood events, not only for the three flood events studied in the article.

Third, there are a lot of language problems with the manuscript. A serious editing by a language proofing company or by a native speaker is warranted to remedy those problems. Below are some specific comments.
Reply: Thank the reviewer for this comments, editing from native speakers has been done.

Specific comments: Lines 36-40: There should be periods between sentences.
Reply: done.

Lines 42-48: these sentences here suggest that hydrological forecasting based on observed rainfall only has a limited lead time. I agree with this point, but for large basins (e.g., the Three Gorges Basin), the forecasting skill gained from hydrological routing based on upstream and downstream streamflow information can be very significant (up to 72 hours or more), especially after major storms. This skill will of course diminish after several days. Therefore, I think the sentences should be modified to reflect this point.
Reply: Here is to discuss the necessity of extending flood forecasting lead time, generally the statement is good, so the authors do not make change to this statement.

Lines 98-101: I think parallel computing has made distributed hydrological modeling less computationally demanding. But to say computational burden does not exist is an over-stretch of the fact. We still face challenges when running hyper-resolution distributed models.
Reply: revision made.

Lines 101-06: on automatic calibration, I don't think there is a well-established way to calibrate distributed hydrological models, even though there are plenty of attempts to do it. I think more careful wording should be used here.
Reply: The authors think this statement is true, so no revision.

Lines 108-122: I think this whole paragraph jumps into conclusion by presenting the results first. If you indicate WRF QPF over-estimates the precipitation, wait until you present the results of this numerical experiment. By the same token, if you suggest that post processing of WRF QPF helps, you also need to present the results first. At this point in your paper, you also need to discuss what scientific problem you intend to solve, what is unique about your approach.
Reply: This is in the introduction section, so some conclusion is presented first is acceptable, detailed methods and results are shown in other parts.

Line 124: "studied area" should be "Study area"
Reply: revised.

Line 138: "focus" -> "focuses"
Reply: revised.

Lines 139-141: the number of events is too small to prove the effectiveness of this approach. A much large sample should be used. It is especially true for the post-processing to be presented later.

Reply: The authors agree that with more flood event, the result will be more robust. But the data is limited in this study.

Line 163: what is "hiemal"?

Reply: it is a typo, revised to winter.

Line 191: the "Li et al. (2015)" paper is not shown in the reference list.

Reply: added.

Lines 206-211: there is no discussion on the WRF model setup in terms of spatial domain definition. What exactly is the grid domain used for the study? How are the lateral boundary conditions or the initial conditions set? How can you justify that 20x20 sq km grid is adequate for the LRB region? Normally people utilize a nested domain approach in order to better capture the local features of the basin. Why this approach is not used here?

Reply: The whole section 3 has been rewrote, and this issue is briefly addressed in the revised article. But as the WRF results were adopted from the reference, so detailed introduction should be referred to the reference(Li et al., 2015)..

Section 3.4: I understand it is necessary to correct the biases in WRF QPF. But how can this correction be applied in real-time forecasting? As shown in section 3.3, the biases change from storms to storms and for different lead times. It is impossible to perform correction consistently because the correction factors change all the time and cannot be determined a priori. I think post-processing is necessary. But the correction factor must be determined on the basis of a large sample, not an individual event. Usually post-processing can be conducted by making use of a big archive of long hindcasts (preferably 20 years or more) using the WRF model under similar setup. I understand it is a huge undertaking and probably cannot be executed by a graduate student.

Reply: In this study, as the data is limited, so the only one event is used to calibrate the WRF QPF, and the other two are used to validate this method. Results of this article shows the effectiveness of this method, if there is more data, one can use more data to calibrate the results.

Line 264: the numbering of the section should be 4, not 3.

Reply: revised.

Line 289: I find the model parameter optimization done here is problematic. First, model calibration should be done using a larger sample of hydrological events. Otherwise the parameters are not going to be robust under different conditions. In this study, one event is used to calibrate the model parameters. If another event is *used, it is highly likely that different optimal parameters would result. Second, I find the post-processed WRF QPF to be problematic as well for reasons I gave in comments related to section 3.4. In theory, if the hydrological model is reasonable and observed hydrological data are reliable, the calibrated parameters based on this kind of information should be fine when used in real-time forecasting. Here in this study, I am not sure if the authors have articulated clearly why calibrated model parameters based on the use of historical hydrological observations should not be used.*

Reply: In lumped modeling, it is widely practiced to calibrate model parameters using a large number of flood events. But in the physically based distributed model, one flood event is enough to optimize the model parameters, this has been validated in a few recent publications.

Lines 308-312: I am not sure how the downscaling is performed. The authors should explain what is the nearest downscaling method.

Reply: The nearest downscaling method is a standard method, so there is no further explanation added.

Lines 314-351: I don't dispute that the hydrological forecasts using post-processed WRF QPF is better than raw WRF QPF in all measures of performance metrics. As I stated before, it is not possible to determine the correction factor in advance, it is hard to justify that we can use the post-processed WRF QPF in real-time forecasting.

Reply: It could be used, but just as the reviewer said, may be more flood events need to have a better correction factor.

Lines 352-386: Again, I return to the question of how robust the optimized parameters are. Model calibration must use a large sample of data to obtain consistent parameters that work under a variety of conditions.

Reply: In lumped modeling, it is widely practiced to calibrate model parameters using a large number of flood events. But in the physically based distributed model, one flood event is enough to optimize the model parameters, this has been validated in a few recent publications.