

**Editor comment for manuscript hess-2017-363**

**Evaluation of the WRF model with different domain configurations and spin-up time in reproducing a sub-daily extreme rainfall event in Beijing, China**

**by Qi Chu et al.**

Dear Authors,

I have read the referees' comments and your replies. While for most points, your replies are satisfactory, I need some more information from you about a major point both referees raised: The choice of modeling and evaluation domains (comment 2 of referee #1, comment 4 of referee #2). In the following, I will explain my current understanding of what you were doing, and my related conclusions. Before I take my decision about how to proceed, I would like to verify that I understood things correctly:

- You model an extreme rainfall event using WRF set up on 3 nested domains (D01, D02, D03) with D01 being the largest domain in coarsest resolution, D03 the smallest, which covers the area of interest, Beijing region.
- When running the models, D01 is forced by ERA-Interim reanalysis, D02 by D01, D03 by D02
- For analysis/evaluation, you map rainfall from D03 back into the results and onto the grid of the D02 domain.
- Analysis and evaluation is the done with this hybrid data set D02+partlyD03, on the D02 domain, against ground rainfall observations and ERA-interim reanalysis.
- The reasons for doing so is that a) the reference truth is available only in a resolution comparable to the D02 resolution, and b) differences among the models are less obvious in the D03 domain than in the D02 domain

So if this is correct, I have two main concerns:

- If your goal is to evaluate different WRF setups with respect to regional (here: D03 or Beijing-scale) heavy precipitation, then a) the evaluation should be done exactly on this scale and b) they should be compared to reference data with adequate resolution for that scale.
- If your goal is to evaluate different WRF setups with respect to larger-scale (here: D02-scale) precipitation patterns, then the D03 run on its small scale is unnecessary. You could then run WRF on the D02 domain with different configurations (the D02 and the D03 settings with respect to parameterization, grid size etc.) and compare these.

So if my concerns are based on correct understanding, you will either have to

- gather better reference data and repeat the evaluation on the D03 scale, or
- do new model runs on the D02 scale and repeat evaluation on the D02 scale

For either way, I feel that more than the time normally spent on major revisions is necessary. As I still think your study is relevant and interesting, I see three possible ways from here: Either you are confident to finish the required work within, 2-3 months, then the decision can be major revisions. If not, you could withdraw and resubmit when you are ready, or I reject the study with the invitation for resubmission. Please let me know what you prefer.

Yours sincerely,  
Uwe Ehret