Review of NHESS-2023-192

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: Chung-Chieh Wang, Duc Van Nquyen, Thang Van Vu, Pham Thi Thanh Nga, Pi-Yu Chuang, and Kien Ba Truong

Summary: This manuscript is a follow on to another manuscript that evaluates an extreme rainfall event that occurred in central Vietnam in 2018. The focus of this manuscript is on an evaluation of the predictability of the event. The authors present an analysis of a time lagged ensemble and perform ensemble sensitivity analysis. The authors conclude that predictability increased as lead time decreased for the event. Additionally, they find various atmospheric features are important to the predictability of the storm. While the case is interesting, I find the analysis superficial and incomplete. The overarching results is that an event becomes more predictable when it gets close to occurring, which is well known and does not add to the body of literature on ensemble prediction. For these reasons, I inform a decision of *Reject*. Should the editor or other reviewers come to another conclusion, I am happy to review the manuscript again.

Recommendation: Reject

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

1. 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.
    a. The result of predictability increases as lead time decreases is not a new result to the body of literature on meteorological prediction.
    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.
2. 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.
3. 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.
    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.
4. 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.

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

1. Lines 21: typo "predicts"
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.
3. Lines 85-87: Citations are needed to support this statement.
4. Lines 182-184: What version of the GFS?
5. Section 2.1.3: Is there any citation or information about the observational error associated with the gauges in this network?
6. Section 2.1.4: A citation needs to be added for IMERG. Additionally, please let us know what version you used.
7. Line 247: Table 2 does not appear anywhere in the manuscript.
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.
9. Section 3.2: What does "per SD" mean?

References:

https://journals.ametsoc.org/view/journals/atsc/69/11/jas-d-11-0315.1.xml?tab_body=fulltext-display

https://journals.ametsoc.org/view/journals/mwre/144/10/mwr-d-16-0083.1.xml?tab_body=fulltext-display

https://rmets.onlinelibrary.wiley.com/doi/full/10.1002/qj.3367