

Interactive comment on “A Comprehensive Evaluation of the National Water Model (NWM) – Height Above Nearest Drainage (HAND) Flood Mapping Methodology” by J. Michael Johnson et al.

D. Blodgett (Referee)

dblodgett@usgs.gov

Received and published: 24 May 2019

The review comments provided by other reviewers are comprehensive and I agree on practically all points. The discussion below adds additional critique from my perspective. I suggest accepting this manuscript with minor revisions per the author's reconciliation of comments received.

This paper is an overview of an evaluation of the combined skill of the operational National Water Model flow estimates, synthetic rating curve transfer to stage and Height

C1

Above Nearest Drainage mapping technique. It demonstrates that the NWM-HAND system of models is capable of producing semi-realistic maps of (binary – flooded or not) inundated area. Given that this is an evaluation of an entire suite of loosely coupled models, the claim that it is "comprehensive" is a stretch. To be a "comprehensive evaluation", each individual model would need to be considered in isolation and/or errors from each model / data source would need to be controlled for. A better description might be "integrated evaluation" or "combined skill evaluation".

The ancillary investigation of causes of poor evaluation results is a valuable contribution and is well thought-out and documented analysis. I would suggest being more explicit that a rough 2D analysis of peak flooded area was used to generally identify flood inundation problem areas and that the causes are investigated through identification of poor performance in specific parts of the coupled NWM-HAND system. Bring some of the limitations and realistic potential of the analysis up from the discussion into the introduction. Also bring a summary of the issues found up to the abstract and/or introductory section. Do any of the errors you found not align with what you would expect from a modeling system with the formulation of the NWM-HAND system?

The paper includes a brief introduction to the Operational National Water Model application of the WRF-Hydro-NWM software and the synthetic-rating-curve / HAND flood inundation model. The description is lacking discussion of the meteorological forcing data, soil / subsurface data, details of how NWM routing and surface storage parameters are derived, and other important aspects of the model that would affect performance and could be included in such an evaluation. The introduction is also lacking a general overview of the NWM's objectives which could/should be used to temper the expectations and focus the aims of an evaluation. e.g. The NWM is intended to predict impending high-flows for the purposes of flood warning and potential inundated area guidance.

Noting that the NWM-HAND system is not used for official forecasts and is to be considered for guidance only at this stage in its development. Given these kinds of caveats,

C2

the evaluation presented in this manuscript is of great value as it demonstrates that the NWM-HAND system is producing flood inundation products that would be generally useful for the intended purpose. Finally, more information on the nature of the retrospective model run should be provided. Given that the NWM is only calibrated in some locations, to a small set of potential calibration targets, it should not be expected to produce realistic flow volumes. Additionally, the retrospective does not assimilate observed streamflow – which should be noted and taken into account here.

For areal comparisons, "Perennial NHD water bodies" are removed – what about "NHD Area" features that are used to denote double line streams? The water body / area features are used inconsistently in the NHD but at the end of the day, the "NHD Area" features should be considered water as they are used as a mask over top of flowlines and artificial paths.

Given that flooding is a 4D (XYZT) phenomena, using a 2D (XY) evaluation technique seems a bit limited and should be justified. Why was the evaluation not weighted using flood depth? Why was the evaluation not weighted by proximity to observed peak timing? I think there is value in the "max flood extent" evaluation, but its choice, limitations, and utility should to be discussed.

Why is the analysis binned by stream order? Given the morphology (stream density) in different parts of the country, stream order is different for streams that are equivalent on other metrics such as modeled-surface-contributing area.

It strikes me that this paper would benefit from some hypotheses based on the characteristics of the modeling approach. e.g. Given that HAND is not a physically based model in that it does not route flow over the landscape or preserve mass, we would expect small errors in stage to produce large errors in inundated areas in low-relief landscapes.

I wouldn't represent issues caused by misuse of the NHD data model in hydrologic model formulation as "errors in the NHD". While the NHD has numerous anomalous

C3

representations of the network, it is just one source of data that must be considered when constructing a hydrologic model. For example, the decision to initiate a headwater flowline is a best guess of perennial flow and the distinction of a sink or a water body that connects to an intermittent flowline is based on the conditions considered for mapping and not extreme hydrologic scenarios encountered by the NWM. The misuse of NHD as a hard-and-fast geospatial framework for hydrologic modeling should be seen as a general flow in the NWM and not characterized as "errors" in a cartographic dataset.

Interactive comment on Nat. Hazards Earth Syst. Sci. Discuss., <https://doi.org/10.5194/nhess-2019-82>, 2019.

C4