

Interactive comment on “Performance of automated flood inundation mapping methods in a context of flash floods: a comparison of three methods based either on the Height Above Nearest Drainage (HAND) concept, or on 1D/2D shallow water equations” by Nabil Hocini et al.

Paul Bates (Referee)

paul.bates@bristol.ac.uk

Received and published: 6 January 2021

This paper tests the ability of three different automatically built flood inundation mapping methods to predict flood extent and high water marks recorded during three extensive flash floods events in French catchments. It differs from much other work in this area that seeks to benchmark inundation modelling methods by virtue of the large spatial scale over which comparisons are undertaken, the large volume of comparison

C1

data (which comprises many thousands of high water marks) and the even treatment of the different methods. Studies in this area can sometimes be undertaken in only localised areas using limited data which cannot discriminate well between competing approaches and papers can also suffer from a kind of ‘unconscious bias’ towards the researcher’s own model. All these pitfalls are avoided in this paper, and the resulting study is therefore a serious one. The research is well executed and mostly very well presented, and I think could be published in HESS with the correction of the following points.

1. I think there should be a bit more discussion of the limitations of using a steady state approximation to model flash floods. I guess this works ok because the automatic model build splits each catchment into small reaches where it is much more plausible to assume steady state conditions, but it would be nice to hear the justification from the authors.

2. Somewhere in the paper there needs to be a discussion about the limitations of using flood extent as a validation metric in narrow valleys and headwater catchments, especially during catastrophic floods events which are very likely to be valley-filling. In these circumstances it may be easy for models to replicate inundation extent and this metric may not be able to effectively discriminate between competing approaches. I have a suspicion that this effect may explain quite a lot about why the performance of the HAND method varies markedly in space.

3. Line 29. Methods are now starting being developed to estimate unknown bathymetry in large catchments which might be worth mentioning here e.g.

Gleason, C. J., & Smith, L. C. (2014). Toward global mapping of river discharge using satellite images and at-many-stations hydraulic geometry. *Proceedings of the National Academy of Sciences*, 111(13), 4788-4791. <https://www.pnas.org/content/pnas/111/13/4788.full.pdf>

Grimaldi, S., Li, Y., Walker, J. P., & Pauwels, V. R. N. (2018). Ef-

C2

fective Representation of River Geometry in Hydraulic Flood Forecast Models. *Water Resources Research*, 54(2), 1031-1057. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1002/2017WR021765>

Neal, J. C., Odoni, N. A., Trigg, M. A., Freer, J. E., Garcia-Pintado, J., Mason, D. C., et al. (2015). Efficient incorporation of channel cross-section geometry uncertainty into regional and global scale flood inundation models. *Journal of Hydrology*, 529, 169-183.

Brêda, J. P. L. F., Paiva, R. C. D., Bravo, J. M., Passaia, O. A., & Moreira, D. M. (2019). Assimilation of Satellite Altimetry Data for Effective River Bathymetry. *Water Resources Research*, 55(9), 7441-7463. <https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2018WR024010>

4. Line 51. The key point about the paper by Savage et al quoted here is that they found that below particular grid scales the model precision became spurious 'given other uncertainties'. Might be worth editing to include this idea.

5. Line 2014-5. the sentence starting "A conventional Dinf ..." could do with just a bit more explanation to be understood by a more general audience not familiar with these terms.

6. Line 110. There needs to be a bit more discussion about the limitations of HAND. My understanding of the method is that it assumes that: (i) the water level is uniform over the reach and (ii) that all cells with elevation lower than the water level are inundated even if there is no flowpath connection to the channel. See Figure 1 in this paper <https://nhess.copernicus.org/articles/19/2405/2019/nhess-19-2405-2019.pdf>. Extended cross section 1D methods can also suffer from the second of these issues. Both assumptions are obviously very different to how floods behave in reality and will explain some of the misprediction with HAND and the 1D model. 2D approaches automatically avoid both issues.

7. Line 117. There are now a few papers on the importance of cross section spacing

C3

in 1D models which you should probably cite here. Would also be worth a sentence discussing how your model build dealt with this issue. See for example: Anuar Md Ali, Giuliano Di Baldassarre & Dimitri P. Solomatine (2015) Testing different cross-section spacing in 1D hydraulic modelling: a case study on Johor River, Malaysia, *Hydrological Sciences Journal*, 60:2, 351-360, DOI: 10.1080/02626667.2014.889297.

8. Line 141. I think this statement needs a reference.

9. Section 3.1. This section needs to include a more extensive discussion of the uncertainties in the observed data. This then needs to be picked up in the discussion to determine whether the models can match the observed data to within error or not. You already discuss the terrain data error in the paper, but don't really say much about errors in the observed discharge other than the rainfall-runoff model generally matched the observed discharge to within 10%. However, discharge gauging during extreme floods is fraught with difficulty and you need to consider the likely error in this, even if this can only be a best estimate made with reference to other studies. The error in the rainfall-runoff model is somewhat misleading, as most such models usually have enough degrees of freedom to be able to match 'observed' data adequately, even if it has error and is disinformative. I would expect discharge gauging during flash floods to be have errors of at least +/-20%. Similarly, you need to say a lot more about how the high water marks were collected, what their likely error are and what QA/QC procedures you undertook to clean up these data. I think for each catchment you should include a plot of HWMs versus thalweg distance and also plot on this the overall valley slope derived from LiDAR data. This will show if there are obvious outliers as we would expect flood water surface profiles to decrease slowly and monotonically in a downstream direction. Regions of supercritical flow may be an exception to this rule of thumb, but, in general, this is the pattern we might expect. Plotting the model water surface slopes on these graphs would also be very informative. A further quality check is to plot the difference in elevation between pairs of high water marks and compare this to the valley slope. Figure 3 in this paper might be a useful template:

C4

Fewtrell, T.J., Neal, J.C., Bates, P.D. and Harrison, P.J. (2011). Geometric and structural river channel complexity and the prediction of urban inundation. *Hydrological Processes*, 25, 3173-3186. (10.1002/hyp.8035).

Lastly, you need to say a lot more about how the inundation extent was mapped and what were the likely errors in this.

10. Section 3.2. It would be good to include a plot of the hydrographs so readers can better appreciate the event dynamics. Did the events lead to any regions of supercritical flow and, if so, how well do you expect the models to perform at these locations?

11. Line 238. This threshold of 65% is arbitrary. I don't disagree with it, but I think you need to better justify its choice.

12. Line 241: What did you do about calculating the error at high water marks where the model did not predict any inundation?

13. Line 250. I tried to download the data but the site did not allow me to do this. Please can you fix this.

14. Section 4: Through this section and the discussion you need to take into account the impact of the observed data errors on your ability to make inferences and discriminate between the three models. Where do the models match the data to within error and where do they not?

15. Line 334. Was there any supercritical flow at the bridge blockages?

16. Line 373. Here you might want to return to the point about the push to finer resolution becoming spurious at particular scales given other uncertainties. See the excellent paper by my co-reviewer Francesco Dottori that first identified this issue:

Dottori, F., Di Baldassarre, G., and Todini, E. (2013), Detailed data is welcome, but with a pinch of salt: Accuracy, precision, and uncertainty in flood inundation modeling, *Water Resour. Res.*, 49, 6079– 6085, doi:10.1002/wrcr.20406.

C5

17. Line 378. One solution to the Near Real Time prediction problem with 2D models is to pre-compute a library of inundation simulations at different flow rates which can then be sampled extremely quickly based on the observed or forecast discharge. The advantage of this approach is that one can sample a range of flows of weight them according to their likelihood, thereby also accounting either for the uncertainty in the observed data or in the forecast. See the following paper by David Leedal for some ideas on how to do this:

Leedal, D., Neal, J., Beven, K., Younger, P. and Bates, P. (2010). Visualisation approaches for communicating real-time flood forecasting level and inundation information. *Journal of Flood Risk Management*, 3 (2), 140-150. (10.1111/j.1753-318X.2010.01063.x).

I hope these comments are useful and I very much look forward to seeing the paper in print.

Paul Bates

University of Bristol

Interactive comment on *Hydrol. Earth Syst. Sci. Discuss.*, <https://doi.org/10.5194/hess-2020-597>, 2020.

C6