Natural Hazards and Earth System Sciences Discussions

Nat. Hazards Earth Syst. Sci. Discuss., referee comment RC1
https://doi.org/10.5194/nhess-2022-263-RC1, 2022
© Author(s) 2022. This work is distributed under
the Creative Commons Attribution 4.0 License.

Comment on nhess-2022-263

Anonymous Referee #1

Referee comment on "Transferability of data-driven models to predict urban pluvial flood water depth in Berlin, Germany" by Omar Seleem et al., Nat. Hazards Earth Syst. Sci. Discuss., https://doi.org/10.5194/nhess-2022-263-RC1, 2022

This paper is concerned with the comparison of two machine learning techniques for predicting urban pluvial flood hazard in areas where the machine learning model was not trained. Namely, these are the UNET approach that has recently been used in a number of studies for the same purpose, and the random forest approach that was similarly used for predicting flooding, but (to my knowledge) not in a context of generating 2D flood maps for specific rain events. In addition, the papers explores the probability of transfer learning for these approaches
The paper is interesting, generally well written and within the scope of the journal. I do have one major criticism and a number of minor comments as detailed below, but I expect that these can be adressed in a revision and that the paper can subsequently be recommended for publication.

Major comments:
The comparison of UNET and random forest (RF) is a (the) key aspect of the paper. The paper concludes that the RF does not transfer well to other areas because it overfits the training data. This effect can and should have been avoided, and since it is such a central part of the paper I do think it needs to be fixed.
The straightforward approach would be to apply a cross validation approach. Divide the training dataset into, for example, 5 areas, perform 4 training iterations where you train on 4 areas and validate on the 5th, select RF hyperparameters that minimize the cross validation loss, and then train with these hyperparameters on the entire dataset.
In particular, this concerns the depth of the decision trees. As I understand from the paper, these have not been limited, and therefore the trees probably simply incorporate the entire training dataset. Please do include results for this in the paper or appendix.

Minor comments:

- Figure 1: The different subareas have quite different properties (models trained on

some areas generalize, while this is not the case for other areas). Please include som detailed illustrations of the subareas (elevation, maybe also flood areas). I think this Figure could become a 4 panel figure, one panel showing the overview of Berlin and the other panels details of the 3 areas. The legend needs to include units and an explanation that it is elevation that is displayed.

- line 92: what kind of storms were fed into the hydrodynamic simulations? block rains? CDS? Euler? ...
- line 173: please provide details on how feature importance was assessed (remove variables and assess relative change of loss function? validation loss or training loss?)
- Figure 4:
  - This figure is impossible to read in black and white print
  - While all the information in the figure is highly relevant, it is also quite convoluted and hard to understand. My guess is that it will be easier to read if you group all UNET results on the left, and all RF results on the right. "UNET" and "RF" could then be moved as headings to the top of the figure, making it a bit easier to recognize the "training domains" text as an axis caption (there is also a typo in the figure). You can then also use the same symbol for RF and UNET results, making the legend consistent with the symbols in the figure.
- Figure 5:
  - I would suggest rearraning this figure in the same way as Fig. 4. Consider also including some more text into the figure, e.g. "baseline training areas" and "transfer test area" so that the figure becomes a bit more self-sufficient. Symbols like SA0->SA1&2 make it very hard to understand the figure without reading everything else in detail.
- Figure 6:
  - Please include a similar figure for a smaller event. The most challenging part for machine learning models is to correctly capture the boundary between flood / no flood. In general, I miss results for how well the models perform for different rain intnsities.

Code (these are not review comments but it would be great if you could adress these):

The datafile on Zenodo is rather big. We tried to download it and failed - it would be great if you could subdivide the data into multiple zip files.

In addition, the modelling scripts start by loading pickle files, but the scripts generating these from the original data are not provided.