Hydrology and Earth System Sciences Discussions

Hydrol. Earth Syst. Sci. Discuss., referee comment RC3
https://doi.org/10.5194/hess-2021-614-RC3, 2022

Comment on hess-2021-614

Anonymous Referee #3

Referee comment on "Deep Learning Methods for Flood Mapping: A Review of Existing Applications and Future Research Directions" by Roberto Bentivoglio et al., Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2021-614-RC3, 2022

This paper performs a review of deep learning approaches applied for flood mapping. In a field that is evolving rapidly, I think this work can make a valuable contribution in ensuring a common understanding of techniques in the community and outlining future research directions.

My concerns are that
-the review is currently not always very precise in distinguishing the contexts in which different approaches are relevant
-it lacks an assessment of which techniques were used over time (some now popular techniques were not available 2 years ago)
-it could occasionally be better at explaining concepts with a focus on the hydrological target group
-generalization of deep learning predictions to locations / events outside the training data is a key aspect that deserves a more prominent place in the paper. Currently this topic is raised in several subsections. It might be useful to provide an overview on what are actually the needs, which can then be used to discuss whether different approaches are conceptually able to meet this (and if this was/was not implemented in current research)
-comparisons of scores across papers need to be interpreted more carefully than what is currently the case. Scores are not necessarily computed in the same manner. In particular, non-flooded areas are not handled consistently in the literature, which has a major impact on the results.

I think all of these issues can be adressed in a revision. I have provided detailed comments below.

Detailed comments

line 51: the automatic discovery of representations is "to some extent" possible. We are still dealing with an input output model. It is quite a common misunderstanding that deep learning can find "any representations", while many relations in hydrology are highly nonlinear and require careful consideration of the data.

line 126-137: I don't think the detailed overview of modelling approaches is needed in this review.
line 190 to 205: this text is somehow misplaced in this section. It is more an assessment of the properties of different techniques and it would probably make more sense to place it after the different layer types were introduced.
Figure 2: I think it would help many readers if the figure illustrates that the convolutional kernels map many pixels to one. Also in the text (line 210), a simply explanation of the kernels (spatially weighted average where the weights are learned during optimization) may be helpful.
Table 2: I believe the correct citation for the work of Guo et al. is 2021, not 2020
Section 3.2.4: The review is generally missing a section that discusses under which conditions a deep learning network can generalize, i.e. predict flooding in different locations
Section 3.2.5: A key issue when assessing flood predictions (inundation and hazard) is the large number of zeros (often >95% of the dataset) which implies that, for example, accuracy scores almost per definition are in the order of 80% and above. This issue needs to be explained here. In addition, binary scores such as CSI are very vulnerable to double penalty issues.
Section 3.4: In general, for flood inundation, it is not completely clear to me whether the authors focus on models that can predict flood inundation (in binary form) given some rainfall or on "gap filling" in remote sensing data. This needs to be checked in all related sections.
Line 472: Due to the 0 problem mentioned above, "slight" increases in accuracy may actually be linked to substantial changes of the quality of a model. The scores therefore need to be interpreted carefully and it is also not guaranteed that all papers computed scores in the same manner.
Line 496: Pham et al. assessed flood conditioning factors, Löwe et al. performed a forward selection to identify relevant topographic variables, Zahura et al. tested feature importance in their random forest model
Table 5: It is not clear to me why not all the papers performing hazard predictions where included in this table? In addition, the error scores may not be comparable across papers (0 problem or similar) which should be mentioned. Also speed up is a difficult quantity to compare, because it depends on the assumed number of numerical simulations that should be performed (e.g. if we assume that we have to assess flood hazard for 1000 rain events, then the speed up factor obtained by a neural network will be much higher than when only 10 events are considered). Most certainly, these assumptions are also not the same across papers and therefore not comparable.
Section 4.2: The discussion on generalization abilities needs to be differentiated a bit more. Both Guo et al. 2021 and Löwe et al. 2021 consider terrain characteristics as input to their models, and in Löwe et al. 2021 generating predictions outside the training dataset was explicitly the focus of the work. As mentioned by the authors, these approaches are in their infancy and have been tested on limited datasets, but these approaches do consider effects of e.g. the built environment in the form of 2D grids.
Section 5.1: While investigating the possibility to consider mesh-based deep learning setups is an interesting direction, the authors present no argument why this should work better than convolutional approaches (which are also used for simulating fluid movements). Other than stated around line 610, they are simply a different data representation with advantages and disadvantages (mesh generation) and may or may not improve performance.
Line 648: From here on the text no longer focuses on meshes (which is the Section heading) but on physical conditioning.
Line 656: I think a formulation that will be easier to understand for many readers is that the PINN can only be trained for a specific boundary condition (such as a specific rain event) and it is subsequently only able to simulate this specific event.
Line 656: FNOs need to be mentioned as one approach amongst many. DeepONets are a widely known alternative and new approaches are constantly developed. The same is true

for DGP in the following section.
Section 5.3: I don't see how GANs fix data scarcity issues (line 680). They are indeed an interesting approach for e.g. gap filling or the generation of rainfall scenarios, but they do not be trained and do not relieve us of the problem that e.g. flood observations are hardly available. The discussion in the first parts of this section goes in a very different direction than the transfer learning approaches (which focus on training models with few data), which creates confusion.

Conclusions
First bullet - this conclusion could be more clear about the methodological preferences being the current status which is developing rapidly.
Line 724 - I would say DL for hazard mapping so far relies on numerical simulations, this may change.
Line 731-736 - Some of the existing architectures do enable generalization but this certainly requires more research and testing. Meshes are one way forward amongst others.
Line 737-741 - Physics-informed learning is not only relevant in a warning context but for virtually all kinds of flood simulations. FNOs and DGPs are potentially interesting approaches, but there are others. Your are overstating the ability of geometric DL which (to my knowledge) has not been tested in the flood context.
Line 742-745 - As mentioned before, there is some logic here that does not make sense, because the GANs need to be trained against observed data. Once we have a GAN, what would be the point of training another deep learning model that only learns to emulate the output of the GAN?

References:
Löwe, R., Böhm, J., Jensen, D. G., Leandro, J., & Rasmussen, S. H. (2021). U-FLOOD – topographic deep learning for predicting urban pluvial flood water depth. Journal of Hydrology, 603, 126898. https://doi.org/10.1016/j.jhydrol.2021.126898
Pham, B. T., Luu, C., Phong, T. Van, Trinh, P. T., Shirzadi, A., Renoud, S., Asadi, S., Le, H. Van, von Meding, J., & Clague, J. J. (2020). Can deep learning algorithms outperform benchmark machine learning algorithms in flood susceptibility modeling? Journal of Hydrology, 592(July 2020), 125615. https://doi.org/10.1016/j.jhydrol.2020.125615
Zahura, F. T., Goodall, J. L., Sadler, J. M., Shen, Y., Morsy, M. M., & Behl, M. (2020). Training machine learning surrogate models from a high-fidelity physics-based model: Application for real-time street-scale flood prediction in an urban coastal community. Water Resources Research, 56(10), e2019WR027038. https://doi.org/10.1029/2019WR027038