Paper ID: hess-2021-596 (A deep learning technique-based data-driven model for accurate and rapid flood prediction)

Response to Reviewer 2

Comment 1

"*The author used the rainfall as the input to generate simulated flood inundation maps. The paper is well organized, and the LSTM model and Bayesian optimization method appears to be correct and effective. However, there are three major issues. First, the summary of the earlier work needed improvement. There are many related papers using data-driven approach to generate flood maps, however the authors do not include them in the introduction. Second, a research may be regarded as a novel study if it resolves a problem or constraint in earlier studies. However, the LSTM is a new neural network layer that can perform better than ANN or linear regression models, and this manuscript does not appear to have demonstrated its novelty. Lastly, this manuscript lacked baseline models and results, which prevented me from knowing how much better the LSTM model is than a simple average baseline, linear regression model, or an ANN model.*"

Response: We greatly thank the reviewer for the valuable comments to improve the paper. We have significantly revised the paper following the suggestions and better explained the novelty of the work. Overall, the paper has been improved in the following aspects:

(1) We have **improved the summary of the previous work** and added more related papers and suggested references to better explain the novelty of the proposed methodology.

(2) We have **enhanced the prediction effects of the LSTM network in predicting not only the maximum water depths, but also flood time series**, which was a constraint/has not been reported in the previous literature. The proposed method enabled the flood prediction in time dimension, which is a

new contribution to the field.

(3) The revised paper adopted the **transfer learning technology to improve the compatibility and generalization ability** of the proposed model. New results showed the model can be well applied to other case studies. The practical application prospect of the proposed method is enhanced.

(4) We have **analyzed the role of Bayesian Optimization Algorithm in the LSTM network**, and the determination of best design scheme of the network. This has not been explored in the previous literature.

(5) Following the reviewer's suggestions, **we have added two flood prediction algorithms (ANN and CNN) as the baseline models to confirm the effectiveness** of the proposed method. New results show that the LSTM is more competitive than the other two algorithms and performed better in terms of evaluation indicators.

Comment 2

"Method section 2.2. If my understanding is correct, all the flood maps are simulated by your physically-based model. Thus, your are developing a deep learning model as a surrogate model of your Mike series models. Such studies have been studied in the past several years using Deep Learning models (see the following papers). If the main difference between your study and theirs is the use of LSTM other than a fully connect layer, this is not novel enough.

Berkhahn, S., Fuchs, L., & Neuweiler, I. (2019). An ensemble neural network model for real-time prediction of urban floods. Journal of hydrology, 575, 743-754.

Lin, Q., Leandro, J., Wu, W., Bhola, P., & Disse, M. (2020). Prediction of maximum flood inundation extents with resilient backpropagation neural network: case study of Kulmbach. Frontiers in Earth Science, 8, 332."

Response: First of all, sorry for insufficiently addressing the novelty of the proposed method in the original version. As responded to Comment 1 that we have significantly revised the paper (including the introduction, methodology and related results) to better explain the contribution of our work. One of the novelties is to predict flood maps in time dimensions (Figure 1 below). Also the Bayesian optimization approach has significantly improved the accuracy of prediction.

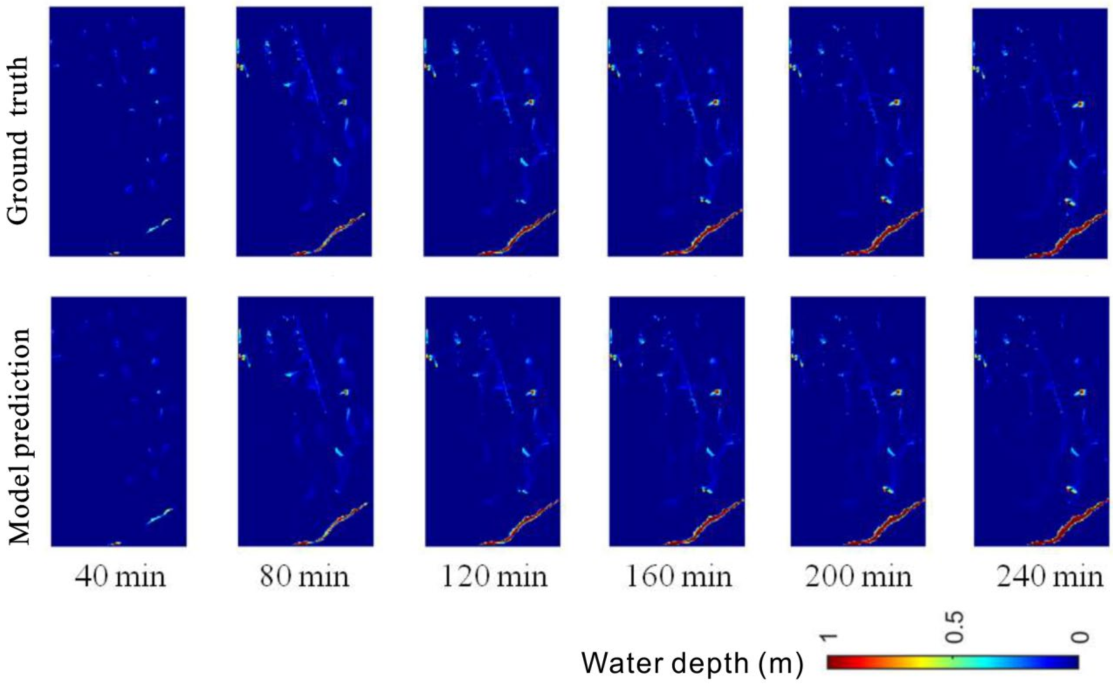

Figure 1: Sample comparisons of flood maps between ground truths and model predictions at different time steps in the entire case study.

Secondly, this study is aimed to develop a flood prediction model that is applicable to various types of case studies, not just as a surrogate model of the physical model. The physical model was used to provide training samples for our model to learn the flood feature extraction ability. In the revised version we used the transfer learning (TL) technology to transfer the LSTM network obtained from the current site (Site A) to another site (Site B), to expand the compatibility and generalization ability of our method. A new case study (Figure 2) was used to test the performance of the method. Results (Figure 3) showed that with TL the proposed model can be well adapted to other cases and the predicted flood maps were consistent to the ground truths. More specifically, the performance indicators of the model in the new case are shown in Table 1. It is shown that the predicted results are very satisfying (BD is close to 0, HID, SS and 2D-CC are close to 1), which proves that flood disaster prediction of the new site can be realized through the TL technology.

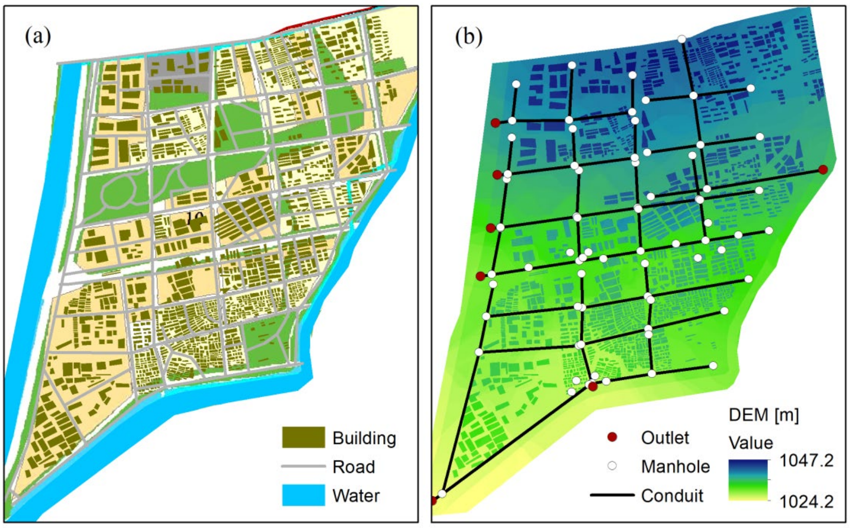

Figure 2: (a) main landuse and (b) drainage network and DEM of new case study.

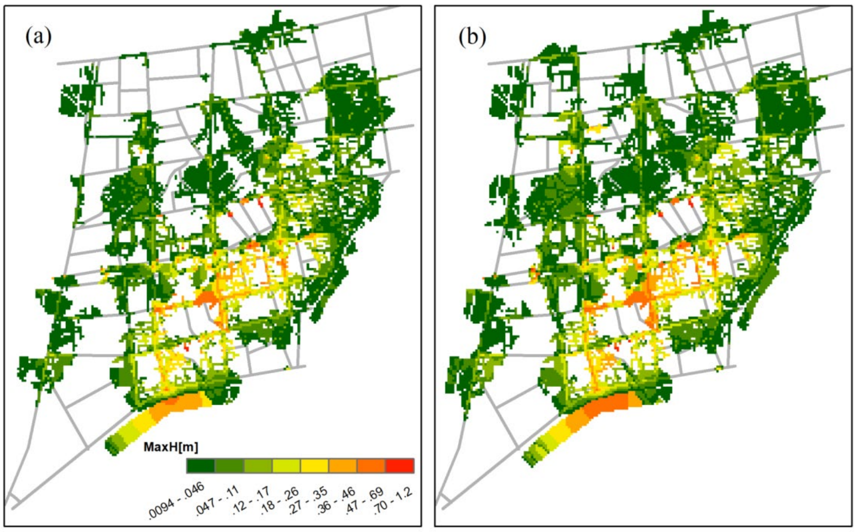

Figure 3: Sample comparison of flood maps of (a) ground truth and (b) model prediction in the new case study under a 50-yr event.

Table 1: The performance indicators of all tested rainfall events in the new case study.

Rainfall events	Performance indicators			
	BD	HID	SS	2D-CC
A	0.003167	0.9994	0.99981	0.997707
B	0.003961	0.999227	0.99995	0.999361
C	0.010744	0.99613	0.997472	0.929869
D	0.003279	0.99948	0.999982	0.999637
E	0.009604	0.99651	0.997349	0.927005
F	0.003337	0.999381	0.99996	0.999301

Comment 3

"What is the color in Figure 6a represents? Can you provide more details about this figure? It seems like the increase of the number of optimizations does not decrease the error much."

Response: We thank this comment to improve the presentation of the optimization results. The color of the bubble chart in Figure 6a represents the size of the error value. We have revised the chart to better demonstrate the variations of the relative errors along with the iterations.

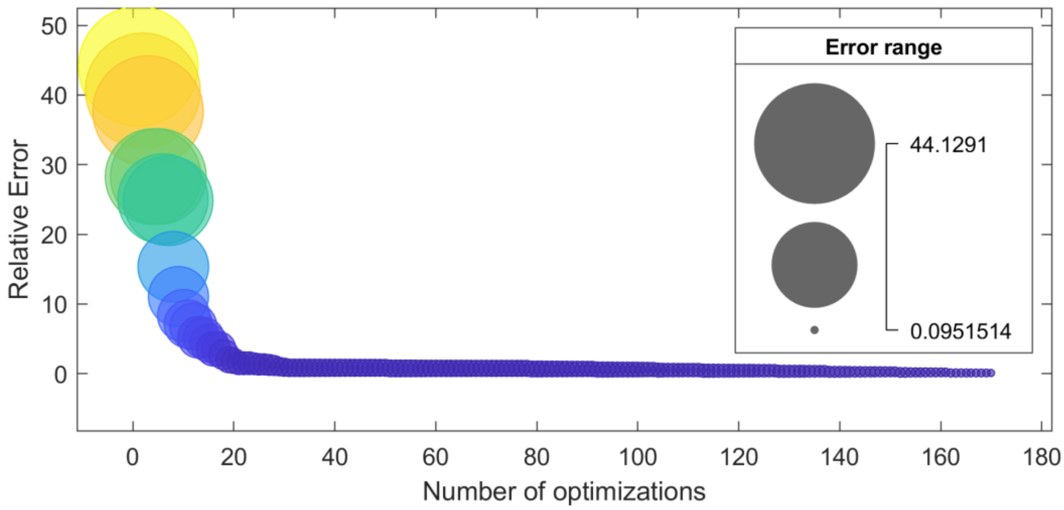

Figure 6: The mean relative errors along with the Bayesian optimization process.

Comment 4

"Are your figures 9 and 10 captions correct? And, is your legend correct for Figure 10? The base color Cyan should represents 0 on your Y-axis, but the legend shows it is 0.5 relative error."

Response: We thank the reviewer for pointing out the problems. Both figure captions were typos and we have corrected them in the revised version. The data of Figure 10 was correct, but the base color was misleading in the previous version because we adopted the 'shading interpolation' function to visualize the figure. We have revised Figure 10 to better present the information and added a subplot to summarize the data for different case scenarios.

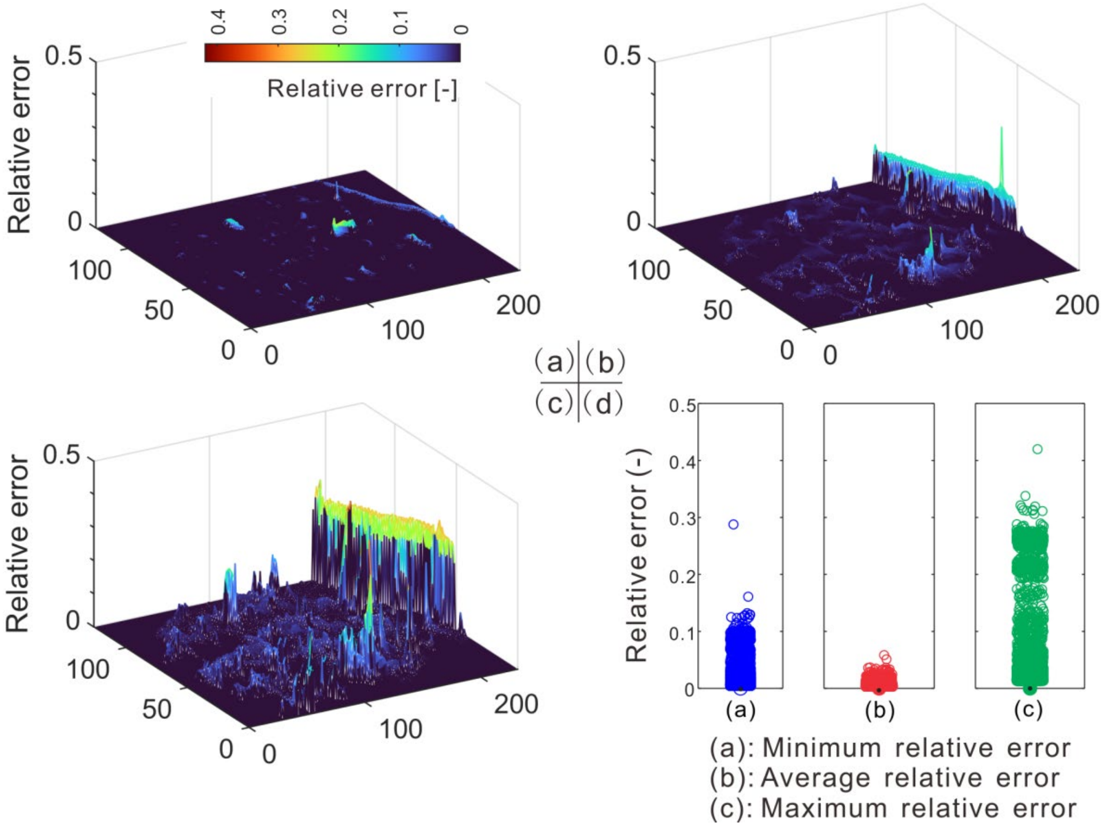

Figure 10: Relative error of (a) best (i.e., minimum), (b) mean, and (c) worst (i.e., maximum) case scenarios, respectively. (d) summarizes the data in boxplots for the three types of scenarios.

Comment 5

"Can you provide results from several baseline models to justify your model performance is good? Some sample baselines could be: 1, models such as ANN as Berkhahn, S., Fuchs, L., & Neuweiler, I. (2019) did (deep learning model using only FC layers other than LSTM). 2, a Lasso or Ridge Regression (or machine learning models) for each point with the overall rainfall as input, water depth as output. 3, an average/median flood map of the training dataset (a.k.a. simple average, see the link below). Without these baselines, your results in Figure 8a and 8b cannot prove much -- we know your model is good, but we don't know how good your model comparing to other simple linear models or simple average of training sets.

https://otexts.com/fpp2/simple-methods.htm"

Response: Following your suggestions, we have investigated two additional flood models (ANN and CNN) to justify the performance of our model. We will explain more details on the ANN (a fully connected network (Back-Propagation neural network)) and CNN (including two convolution layers, one pooling layer and a fully connection layer) models in the revised literature review and Methodology section.

The new results (Figure 4) showed that the LSTM model outperformed the two baseline models in terms of evaluation indicators on both the relative error and the degree of similarity. This confirms the excellent performance of LSTM in flood prediction on water depth and spatial distribution. The ANN performed poorly in predicting water depths and there were a large number of cells associated with large errors. Regarding the BD, HID and SS, the CNN was the least ideal in predicting the spatial distributions. One possible reason could be that the convolution operation of CNN filtered part of the feature information of flood distribution. Note that the ANN's prediction based on the 2D-CC indicator was worst. This could be due to that the fully connected network structure of ANN was prone to overfitting, and may also be interfered by some redundant information.

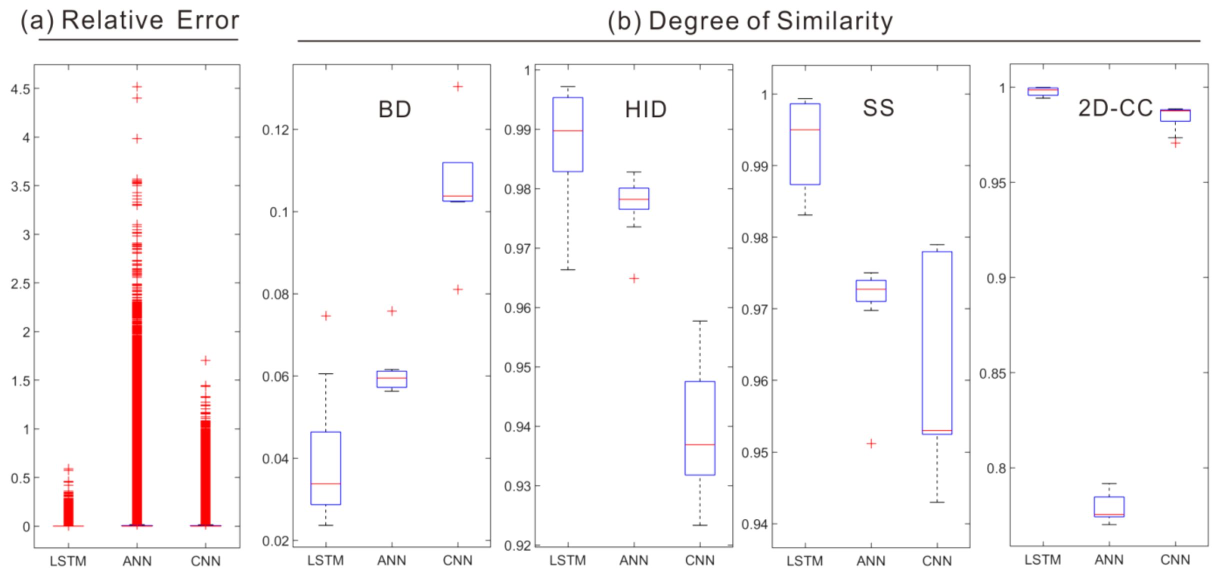

Figure 4: (a) the mean relative error and (b) degree of similarity indicators of the proposed LSTM and two baseline models (ANN and CNN), respectively.

Furthermore, a sample illustration of the predicted flood maps by the three types of models is shown in Figure 5. It is clear that our proposed model is more competitive in flood predictions than the other two methods.

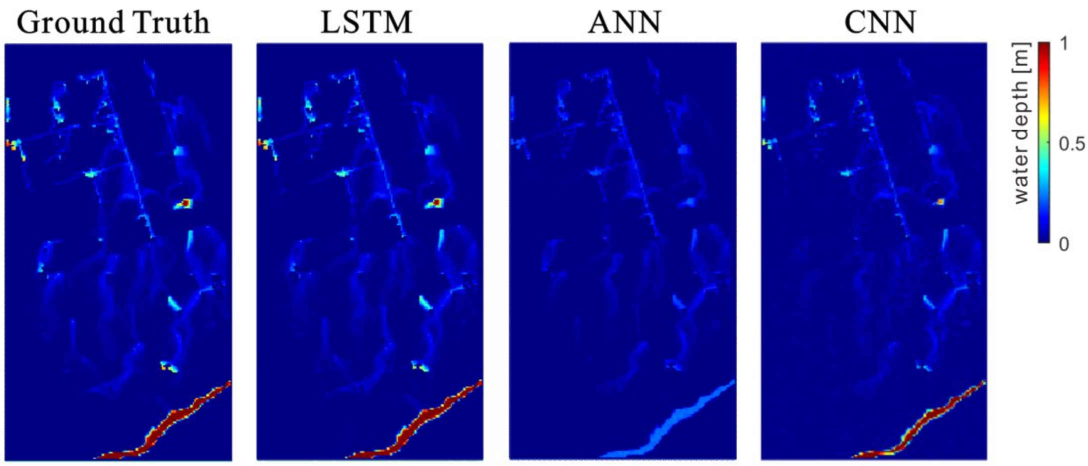

Figure 5: A sample comparison of flood inundation maps of ground truth, LSTM, ANN and CNN models under an 85-year rainfall event.