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

Response to Reviewer 1

Comment 1

"*I find the topic of the submitted manuscript very interesting and also within the scope of HESS. Rapid flood prediction mapping in urban areas has a few challenges when compared with flood modelling purposes (e.g., fluvial flood mapping). The authors identify these challenges and propose a data-driven flood model based on LSTM networks and Bayesian optimisation.*"

Response: We greatly thank the helpful and encouraging comments from the reviewer and the acknowledgment on the contribution and value of our work. We have fully addressed each of the reviewer's comments in the revised version and the paper has been greatly improved thanks to the reviewer.

Comment 2

"The optimisation part is interesting, but the implementation and justification of the LSTM flood model lacks, in my opinion, novelty. One of the arguments of the authors for using LSTM to predict flooding is its suitability to predict time series, i.e. to include the time dimension in the flood prediction mapping. However, they fail to do so as they predict only the maximum water depth maps. This has been presented in previous recent studies (that are correctly acknowledged by the authors); so, what are the novel aspects of this study? Simple a different network architecture? I believe this is not for a scientific contribution that aims to contribute to the advance of the (applied) science."

Response: We agree with the reviewer and have greatly enhanced the methodology and related descriptions and results. To improve the novelty of the paper, in the revised version we have revealed the advantages of the LSTM network on flood time series predictions. Specifically, the following contents have been incorporated:

(1) Create an additional dataset for flood time series. To research the prediction performance of LSTM on flood time series, an additional dataset including 11 rainfall events and resulted flood time series has been included. The flooded water depths were recorded every ten minutes for the entire case study under each rainfall. Among that, nine rainfalls were used for training and the other two rains for testing.

(2) Explore the LSTM network for time series predictions. With the revised methodology and dataset, the tested results showed that our model can well predict flood variations at different time steps. Figure 1 below shows the sample comparisons between ground truths and model predictions of flood maps in time dimension. In visual quality, the predicted flood maps were in good agreement with ground truths at all time steps. The overall prediction effects (based on the relative error) and the evaluation indicators on the degree of similarity were summarized in Figure 2 and Figure 3. These indicators further validated the model performance, which was also satisfying in predicting flood time series.

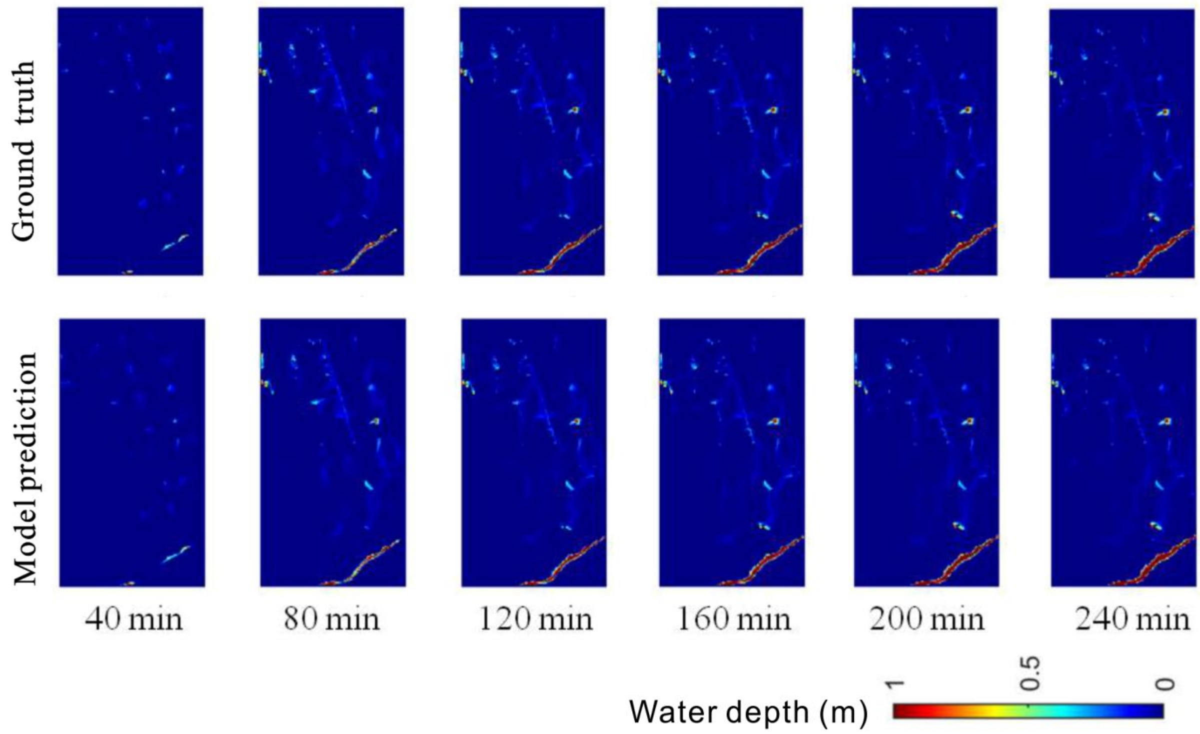

Figure 1: Sample comparisons of flood maps between ground truths and model predictions at different time steps in the entire case study.

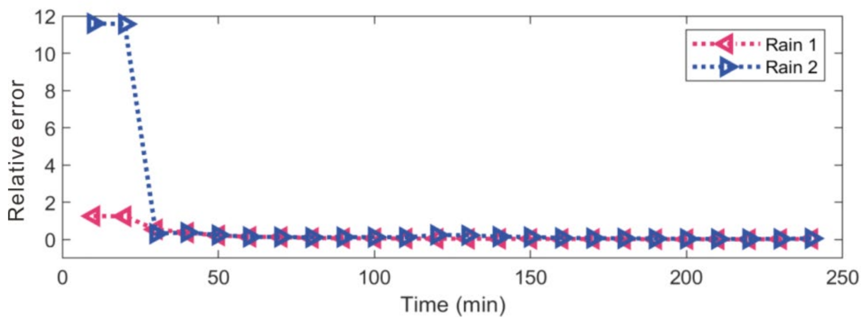

Figure 2: Relative error of flood predictions at different time steps.

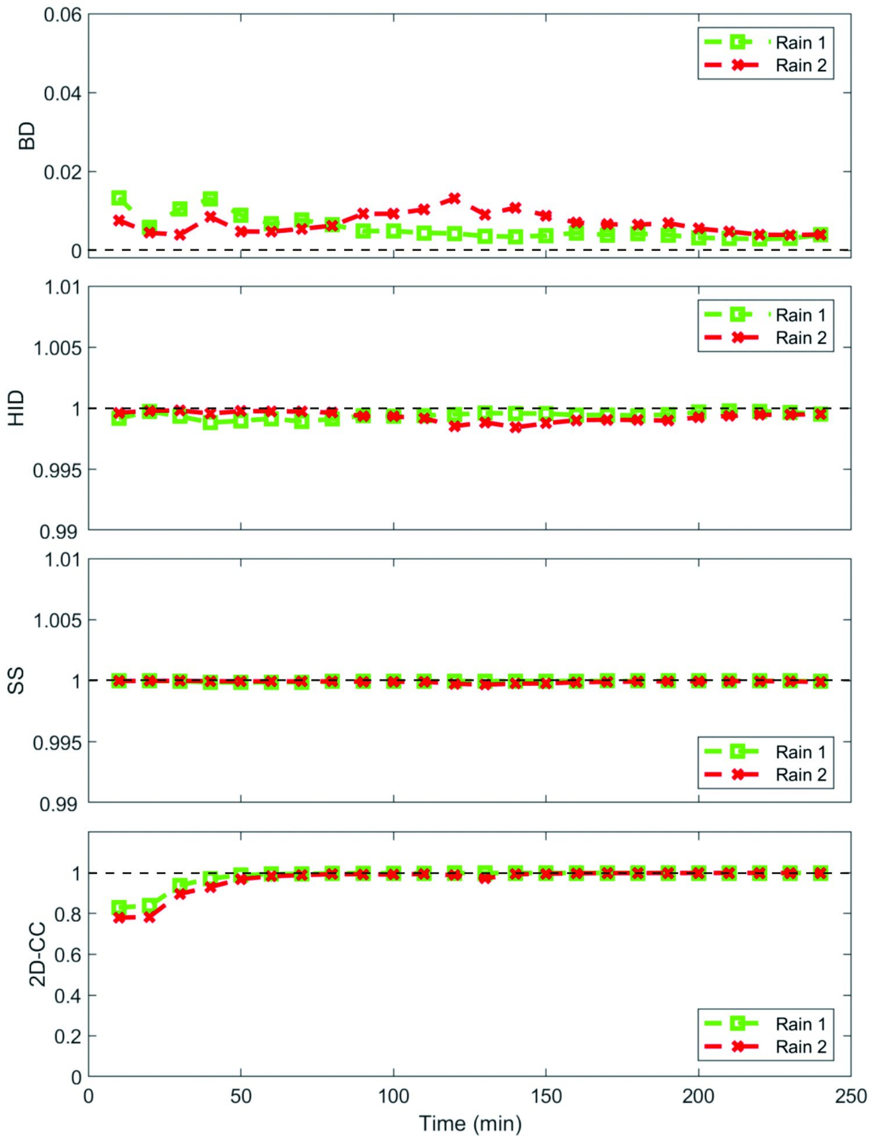

Figure 3: Degree of similarity (BD, HID, SS and 2D-CC) of the flood predictions of testing rainfalls at different time steps. Specifically, the closer the BD is to 0, and the HID, SS, 2D-CC are to 1, the better the prediction performance.

Comment 3

"One of the main challenges about data-driven models, in general, is the capability of the models to generalize to different case studies or contexts. This aspect is not investigated nor discussed in the manuscript - it is only briefly mentioned, and for the 1st time, in the conclusions section. Since terrain elevation is not part of the input data set, it seems that the proposed model is not at all generalisable to other cases."

Response: We agree with the reviewer that one of the main challenges of data-driven models is their compatibility. In the field of machine learning, the transfer learning (TL) technology can significantly improve the application field of intelligent algorithms. In this paper, **we used transfer learning technology to transfer the LSTM network obtained from the current site (Site A) to another site (Site B), so as to expand the compatibility and generalization ability of our method.**

The TL is a deep learning method to transfer knowledge from one domain (source domain) to another domain (target domain), see Figure 4. Through the training of a source model (pre-trained network) using the source data (Site A), the pre-trained network can gain a strong ability of feature extraction. Subsequently, with fine-tuning (transfer learning) the new data (Site B), the pre-trained network can quickly adapt to the new site under different scenarios. With this method, a lot of training time can be saved for the target domain (the new site), and better training effects can be achieved, especially when there are limited training samples in the target domain.

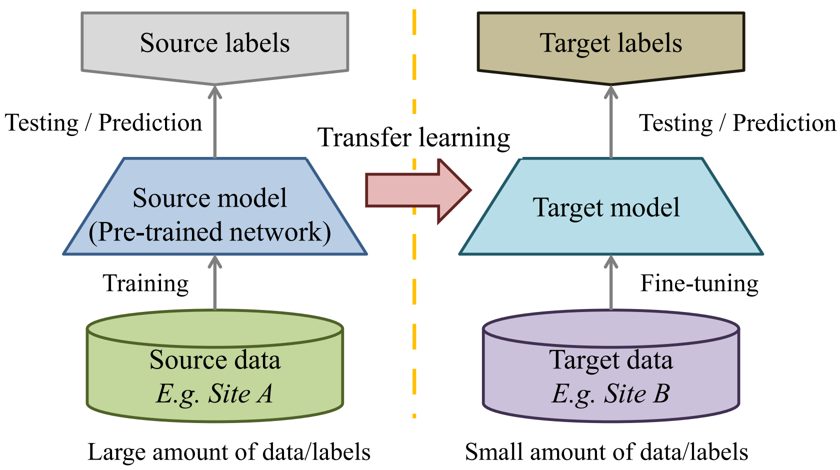

Figure 4: Transfer learning technology.

In the revised version, we have included a new case study to test the performance of our method with the TL technology. The main landuse, drainage system and surface elevation are shown in Figure 5. The tested results (Figure 6) showed that with TL the proposed model was applicable and generalisable to other cases and **<u>the predicted flood maps were consistent and similar to the ground truths.</u>** Specifically, Table 1 shows the achieved performance indicators of all tested rainfall events. The obtained BD is close to 0, and HID, SS and 2D-CC are close to 1, which means the model predictions are highly similar to the ground truth results. **<u>This further proves that the flood prediction of the new site can be realized through the TL technology</u>**.

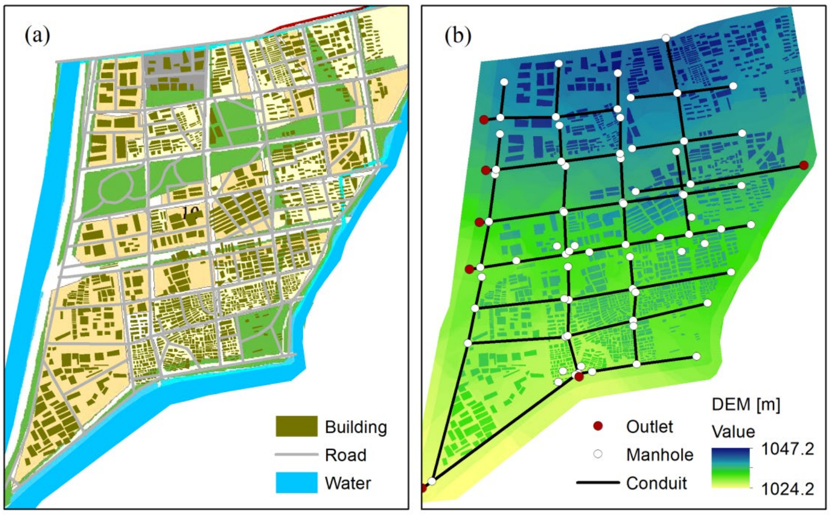

Figure 5: (a) landuse and (b) drainage network and DEM of the new case study.

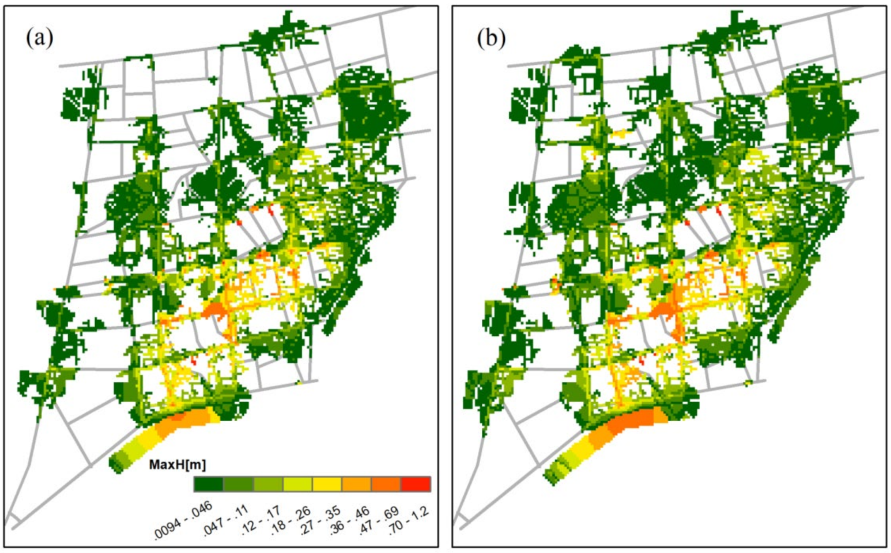

Figure 6: Sample comparison of flood maps of (a) ground truth and (b) model prediction in the new case study under a 50-yr event.

Table 1: The performance indicators of tested rainfalls 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 4

"*The limitations of the model and study presented at the end of the Conclusions section are very similar to those of other previous studies. If the authors are aware of these limitations from previous studies, I would expect them to try to address at least some of the previous studies limitations to improve the knowledge.*"

Response: Thanks for the suggestions. We have improved the methodology and better explained the novelty of the revised paper. We have also better present the related literature and revised the Conclusions section. To summarize, the paper has been improved in three aspects following the reviewer's comments.

(1) Comment 2: enhance the model capability in **predicting flood time series,** highlighting the prediction effects of the LSTM network in time dimension

(2) Comment 3: the adoption of **transfer learning technology** to improve the compatibility and generalization ability of the proposed method

(3) Comment 5: explore the role of **Bayesian Optimization Algorithm** in the LSTM network, and the determination of best design scheme of the network

Comment 5

"*As mentioned above, I think the part of the optimisation could be better explored in the manuscript. Perhaps this could be the novel contribution of the study?*"

<u>Response:</u> Thanks for the suggestions. In the revised version, we have added more methodology descriptions and parameterization analysis of the Bayesian optimization. Specifically, it included the following contents:

(1) **<u>Clarify more optimization process and details in the methodology section,</u>** see the added workflow of the Bayesian optimization process in Figure 7.

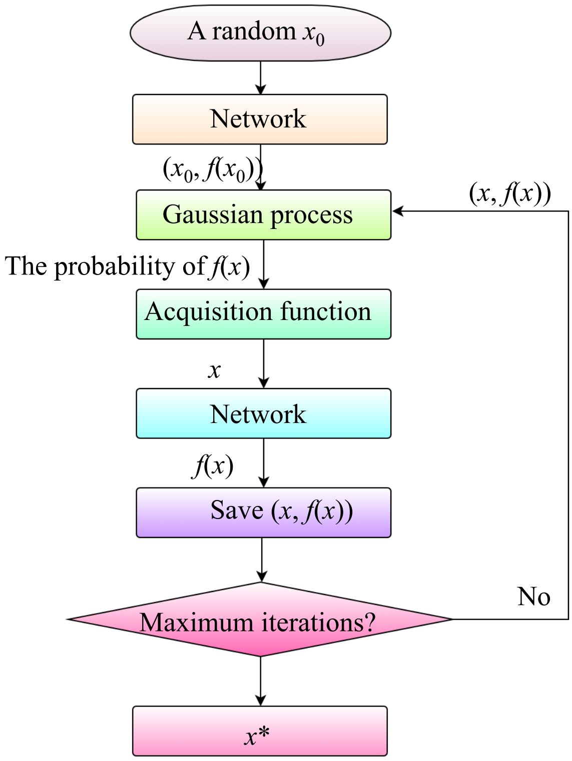

Figure 7: The Bayesian optimization process.

The optimization process is implemented as follows: (1) A group of hyper-parameter combination x_0 (e.g., maxEpochs, learning rates) is randomly selected within the value ranges of the hyper-parameters (shown in Section 2.3.2); (2) The x_0 is input to the network for training to obtain the corresponding objective function $f(x_0)$; (3) The probability distribution of $f(x)$ corresponding to x is calculated and predicted through the Gaussian process using all the inputs $(x, f(x))$; (4) The optimal x is determined by the acquisition function in the probability distribution; (5) The x obtained from Step (4) is taken as the hyper-parameter combination of the network to train, and calculate the objective function $f(x)$; (6) Before reaching the maximum iteration number, the $(x, f(x))$ obtained in Step (5) will be used as the input of the

Gaussian process to continuously update the probability model to obtain a new $(x, f(x))$. Once the maximum iteration number is reached, the x corresponding to the minimum value of $f(x)$ is taken as the optimal hyper-parameter combination x^*.

(2) In the **Result section**, we **analyzed the influence of network parameters** (i.e., network depth, Epoch, MinBatchSize and initial learning rate) **on the prediction results** (see Figure 8).

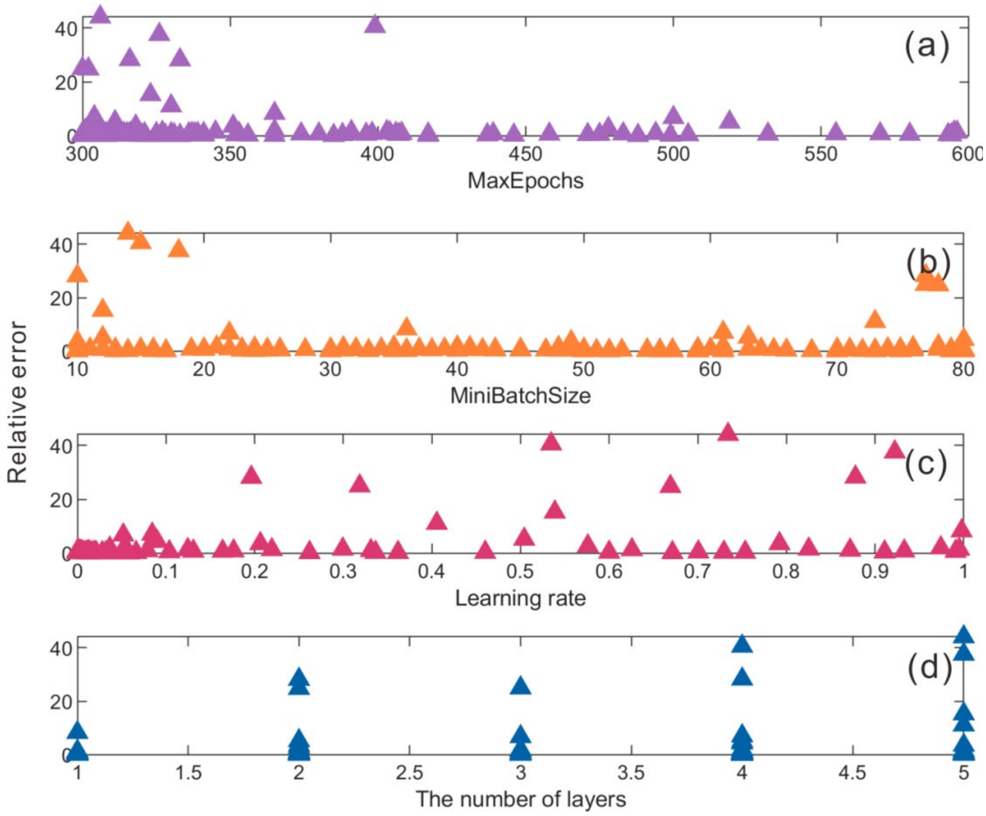

Figure 8: Influence of four types of network parameters on model prediction performance (i.e., relative error).

The results show that: (1) There are large errors when the values of MaxEpochs (i.e., maximum number of epochs) are set too low. Increasing the number of training epochs can avoid adverse events. (2) The MiniBatchSize has little influence on the prediction results, but it is not appropriate to take too large or small values. In this case, the MiniBatchSize of 20-70 can ensure an ideal prediction effect. (3) It is recommended to set a low learning rate. When the value is low, the achieved error is small and close to 0. (4) A deeper network layer can obtain a smaller prediction error.

With the parameterization analysis, the best design scheme of the LSTM can be determined through the optimization.

Comment 6

"*The manuscript is, in general, well written, making it easy to read.*"

Response: We thank the reviewer for the positive feedback and enlightening advices.

Specific comments:

1. "*Line 43: this is valid also for physically-based models. Rephrase?*"

Response: Comments incorporated in the revised version.

2. *"Line 113 & 115: why different models Mike Urban & Flood vs Mike 21? Please provide justification or mention only the model used."*

Response: We agree and have clarified in the revised version that 'Mike Urban' was used as the physical model.

3. *"Line 145: can "... for the long-term memory of data" be better described?"*

Response: Yes, we have replaced "the long-term memory of data" with "the time-varying data".

4. *"Line 149: unclear sentence. It seems that something is missing."*

Response: We agree and have added a new sentence to better explain the information: "It is a priori probability that is used as the basis for selecting the parameter combination in the next iteration".

5. *"Lines 279 - 282: this can't be seen in the plots. The colour scale does not have units. How does the colour scale relate to the yy axis?"*

Response: We have revised the original Figure 9 and Figure 10, and the related legends, captions and texts to better clarify the information and discuss the results. The units of color scales in Figure 9 and 10 are meter (i.e., water depth) and relative error (-), respectively.

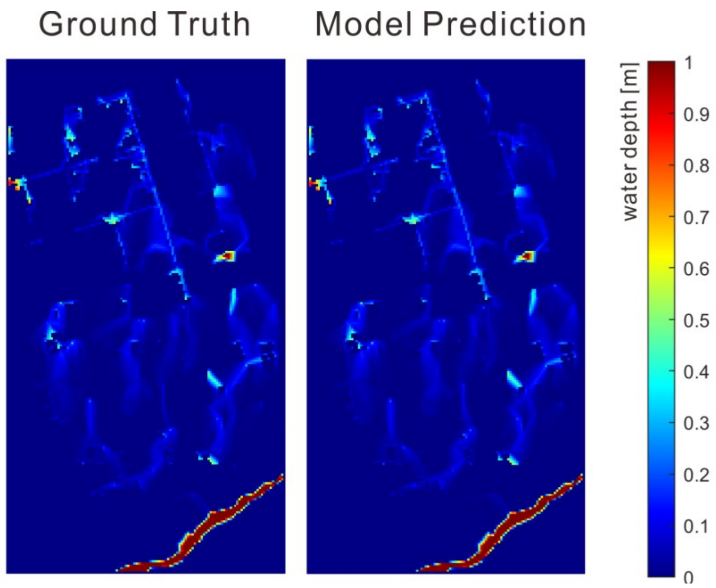

Figure 9: Sample comparison of flood maps between ground truth and model prediction in the best case scenario.

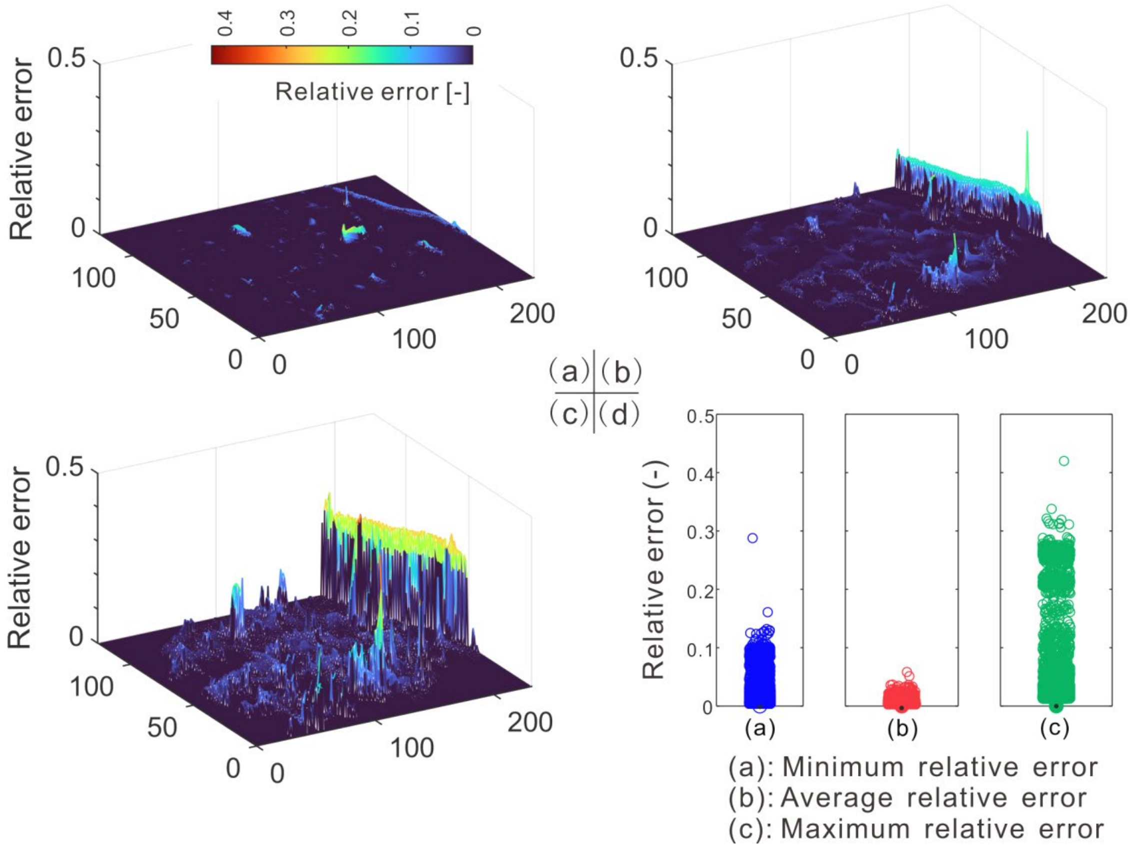

Figure 10: Relative error of (a) best (i.e., minimum), (b) mean, and (c) worst (i.e., maximum) case scenarios, respectively. (d) summarizes the relative error data in boxplots for the three types of scenarios.

6. *“Line 306: the worst and best cases are also interesting to be analyzed and discussed.”*

<u>Response:</u> We agree and have added more data and discussions on the worst and best case scenarios. Figure 10 has been revised to better present the information. It is shown in Figure 10d that in all cases, the mean errors were below 1%. The errors were much higher in the worst case, where there were a small number of cells associated with relative errors greater than 20%. These high-error cells were mainly located in/near the water bodies.