Reviewer 2

It is important and meaningful to improve prediction accuracy in modeling works. Nowadays, application of machine learning including deep learning techniques may be very promising to support conventional modeling approaches. This study constructed LSTM model to simulate surface/subsurface flow and E. coli concentration in a catchment and compared the performance with HSPF model, which is a well-known watershed model. The results are quite interesting and can be useful in scientific and practical fields. I think that this study can be considered as a publication in the journal with **minor revision**.

Some comments are as follows.

We thank the reviewer for finding our work interesting and useful. We have revised the manuscript with reviewer's comments. The answer to each of the comments are given below.

In construction of LSTM in a catchment, this study used only meteorological data as an input to predict flow rates. An issue is that how we can consider characteristics of catchment such as land use and soil property in simulation of flow rate.

Response: We agree with the reviewer that the flow rates in a catchment are affected by the catchment characteristics such as soil characteristics, slope etc. The LSTM is in principle designed to extract temporal features from time-varying input data. The static data consisting of catchment characteristics can however be fed to the along with continuous data. However, in this study we did not consider this because the study consisted of only single catchment. In such a scenario, the LSTM will be trained with only single constant value for each catchment feature. On the other hand, if had data for several catchments, then LSTM could be trained to learn different catchment characteristics. We adopted similar strategy in the preceding study (Abbas et al., 2020) where we trained LSTM with input data of different HRUs and the static input data of HRUs was used along with time series data of each HRU.

I think that the basin area in this study may be relatively small. What if LSTM model application in a largescale watershed? Is the meteorological data enough to predict flow rate? I think that this discussion can be very informative to readers in LSTM application to watershed scale.

Response: The prediction performance of deep learning models is strongly affected by the data distribution of the training data. In order to make a regional or global prediction model, the LSTM should also be trained with input data from more catchments. Some recent work is being carried out in this direction to build regional streamflow prediction models using deep learning. However, due to scarcity of water quality data, building such a regional model is more challenging. Nevertheless, we think the approach adopted in this study can be used as guideline for building regional water quality model by training the model with more input data. We have added this discussion in the manuscript.

Lines 432 – 441: "Deep learning models are based upon the (independent and identically distributed) (IID) assumption which means that the validation data is expected to have the same distribution as that of the training data (Kawaguchi et al., 2017). However, this is not a realistic assumption and it is considered as one of the challenges for researchers in machine learning (Bengio et al., 2021). Thus, in order to build regional or global hydrological models, the deep learning model should be trained on catchment data from diverse catchments. Several researchers have adopted this approach to build regional models for streamflow prediction (Anderson and Radic, 2021; Kratzert et al., 2019; Xiang et al., 2021). However, a similar approach for building regional water quality model will be more challenging due to the scarcity of water quality data.

We hope that the lessons from this study can be used a guideline to train neural networks on regional water quality data".

Line 29, full name is needed for "PDR"

Response: We have corrected this by writing the full name of PDR. The modified sentence is as follow

Line 29: In this study, we simulated the fate and transport of *Escherichia coli* (*E. coli*) in a 0.6 km^2 tropical headwater catchment located in Lao People's Democratic Republic (Lao PDR) using a deep learning model and a process-based model.

Line 52, what is the meaning of "less dangerous than other pathogens"?

Response: We agree with the reviewer that the word 'less dangerous' is vague and not clear. Therefore, we have removed this word from the sentence. The modified sentence in manuscript is as follows;

Line 52 – 53: "*Escherichia coli* (*E. coli*) has been frequently used as an indicator of fecal bacteria because it is easy to culture (Rochelle-Newall et al., 2015)".

In study site description, basin area is needed.

Response: We have added the catchment area in the study site description.

Line 112: "The study area is 0.6 km^2 the Houay Pano headwater catchment".

Line 167, what is the meaning of "rewrote" Did you modify the source code? Rephrase it.

Response: The original HSPF code is in FORTRAN programming language which is difficult to use with modern optimization algorithms and change. Therefore, we converted the code into Python programming language. Thus, we did not modify the source code but converted it into Python programming language.

Lines 170 – 171: For this study, we converted the original FORTRAN code of *E. coli* module of HSPF into Python programming language.

Line 192-193 and 376-378, It is difficult to understand scenario 1 and scenario 2.

Scenario 1 is land use change with same E. coli loading (Fig. S1-a and b) and scenario 2 is land use change with variable E. coli loading in terms of land use (Fig. S1-a and c))? It is confusing.

Response: The purpose of these two scenarios is to assess the impact of different input features. In scenario 1, the land use change information and *E. coli* source information is represented by separate input features. This results in increase in number of input features. In scenario 2, the number of input features was reduced by combining the land-use change information with that of *E. coli* concentration. We have rephrased the sentences to make it clearer.

Line 205 – 208: "In scenario 2, the number of input features was reduced by multiplying *E. coli* source with land-use change. In this way, we calculated *E. coli* source per area for each land use and used this as input instead of using land use and E. coli information as separate input features".

Line 265, among the 10 most sensitive parameters? 10 variables are equally sensitive?

Response: The sensitivity of the 10 parameters is not equal. The sensitivity rank of these parameters is given in Table S2.

Table 2 and line 313, number of optimal batch size and lookback steps are mismatched between the table and sentence. 128 vs 100 and 50 vs 5 h

Response: We thank the reviewer for pointing the mistake. We have corrected the values of batch size and hidden units for LSTM for surface and sub-surface flow estimation. However, the value of lookback steps is correct i.e. 5 hours. This is because the 5 hours of historical data was used as input for both flow estimation E. coli concentration estimation.

Lines 331 – 332: The optimal batch size and LSTM units were 128 and 64, respectively.

Figure 8, what is (a) and (b) in the figure? (a) July 15 and (b) August 1?

Response: In Fig. 8 (a) and (b) represents prediction performance of models during two selected periods with several storm events. The first period is from July 15 to July 23 while the second period is from August 1 to August 5 2017. We have rephrased the caption of Fig. 8 to make this clearer.

Figure 8: *E. coli* concentration of HSPF and LSTM during July 15-22 (a) and August 1-5, 2017 (b). Both storm events were affiliated in validation period.

Line 362 – 370, where is minmax and logarithmic transformation from? There was not any mention about application of minmax and logarithmic transformation in method. All E. coli simulations was based on logarithmic?

Response: We have added information about minmax and logarithmic transformation of the E. coli data in the methods section (2.2.2). We trained the neural network on the transformed data. However, we calculated the performance metrics by transforming the predictions back to normal scale.

Line 198 – 202: "The preprocessing of the data before feeding the neural network can have a significant impact on the of performance (Banhatti and Deka, 2016). Therefore, we compared the performance of model by transforming the *E. coli* concentration using the minmax transformation and the logarithmic transformation. The minmax transformation results in data between 0 and 1 while logarithmic transformation transforms the data on a logarithmic scale".

Line 376-378, Referencing of the figure in the sentences may be wrong. Not Fig. S2 but Fig. S1. Check it..

Response: We are thankful to the reviewer for pointing out this mistake. We have corrected this mistake.

Lines 399 – 400: In scenario 1, we used land-use change time-series information (Fig. S1a) and bacterial source information (Fig. S1b).

References

Abbas, A., Baek, S., Kim, M., Ligaray, M., Ribolzi, O., Silvera, N., Min, J.-H., Boithias, L., and Cho, K. H.: Surface and sub-surface flow estimation at high temporal resolution using deep neural networks, Journal of Hydrology, 590, 125370, 2020.

Anderson, S. and Radic, V.: Evaluation and interpretation of convolutional-recurrent networks for regional hydrological modelling, Hydrology and Earth System Sciences Discussions, 1-43, 2021.

Banhatti, A. G. and Deka, P. C.: Effects of Data Pre-processing on the Prediction Accuracy of Artificial Neural Network Model in Hydrological Time Series, in: Urban Hydrology, Watershed Management and Socio-Economic Aspects, Springer, 265-275, 2016.

Bengio, Y., Lecun, Y., and Hinton, G.: Deep learning for AI, Communications of the ACM, 64, 58-65, 2021.

Kawaguchi, K., Kaelbling, L. P., and Bengio, Y.: Generalization in deep learning, arXiv preprint arXiv:1710.05468, 2017.

Kratzert, F., Klotz, D., Shalev, G., Klambauer, G., Hochreiter, S., and Nearing, G.: Towards learning universal, regional, and local hydrological behaviors via machine learning applied to large-sample datasets, Hydrology and Earth System Sciences, 23, 5089-5110, 2019.

Rochelle-Newall, E., Nguyen, T. M. H., Le, T. P. Q., Sengtaheuanghoung, O., and Ribolzi, O.: A short review of fecal indicator bacteria in tropical aquatic ecosystems: knowledge gaps and future directions, Frontiers in microbiology, 6, 308, 2015.

Xiang, Z., Demir, I., Mantilla, R., and Krajewski, W. F.: A Regional Semi-Distributed Streamflow Model Using Deep Learning, 2021.