Reviewer 1

The authors performed modelling of the transport of Escherichia coli (E. coli) in a tropical headwater catchment located in Lao PDR using a deep learning model and the Hydrological Simulation Program–FORTRAN (HSPF). The deep learning model was built using the long short-term memory (LSTM) technique, whereas the process-based model was constructed using the HSPF. Their results show that the LSTM provided accurate results for surface and subsurface flow, by showing 0.51 and 0.64 of Nash–Sutcliffe Efficiency (NSE), respectively, whereas the NSE values yielded by the HSPF were -0.7 and0.59 for surface and subsurface flow. The simulated E. coli concentration from LSTM also improved, yielding an NSE of 0.35, whereas the HSPF showed an unacceptable performance, with an NSE value of -3.01. The subject is interesting, important and useful. However, there are still some key points need to be addressed. This reviewer recommends to do some revision taking into account the below comments.

We are thankful to the reviewer for his valuable and insightful comments. We have revised the manuscript according to comments of the reviewer. The detailed response to each comment of the reviewer is given below.

We would like to inform the reviewer that the line numbers in our responses correspond to updated manuscript.

Line 62, add g in "determining".

Response: We have removed the typo in line 62 and modified the sentence.

Line 61 – 64: "Thus, they can be a means to determine the fate and transport of fecal pathogenic microorganisms at the catchment scale by simulating *E. coli* in environmental compartments, such as the soil surface and streams (Ligaray et al., 2016; Pachepsky and Shelton, 2011)".

Line65-84, although the authors have performed a good review of literature of process based models, some latest literatures of water quality should be introduced, such as E.coli (Ligaray et al., 2016; Sowah et al., 2020), and limitations of process-based models (Wang et al., 2020).

https://www.sciencedirect.com/science/article/pii/S0048969720341917?casa_token=itdrAAzone8AAAAA:boSZBTjS_8FQY3RZ2bJz9zwQjiwpz9QOLuLvqK2iB6_CMu7NFaqmjvYcaCwPxwvLO7yw1XEQKA

https://www.sciencedirect.com/science/article/pii/S002216941931248X

https://www.sciencedirect.com/science/article/pii/S0048969720326097

https://www.mdpi.com/2073-4441/13/4/518

Response: We thank the reviewer for the literature suggestion. This literature provides important insights about the latest research on *E. coli*. We have included the important highlights from this literature in our manuscript.

Lines 75 - 81: Recently, Sowah et al. (2020) applied the SWAT model to research the sources and drivers of *E. coli* in Clouds Creeks watershed, USA. However, the process-based models still have limitations to accuracy due to the complexity of relationships among hydrological and environmental variables (Abimbola et al., 2020). In addition, the simplified equations of these models can increase the inherent uncertainties, resulting in simulation errors. To overcome these limitations, several modifications of the *E. coli* module of the SWAT model have been proposed to incorporate the impacts of the multiple drivers of *E. coli* fate and transport (Cho et al., 2016; Jeon et al., 2019; Meshesha et al., 2020).

Line85-95, it is unclear what advantages DL has over process-based models. Are there any disadvantages of DL, compared to process-based models?

Response: Deep learning-based models have the advantage over their process-based counterparts because of their high accuracy, faster prediction time and their ability to model complex relationships between input and output features. This ability of deep learning comes mainly from their ability to exploit special form of compositionality in input data by creating abstract features in different layers of neural networks. We have added the following lines to explain more clearly the advantages of deep learning-based models over process-based models.

Line 88 – 92: "Deep learning-based models are superior to their process-based counterparts due to their high accuracy, faster prediction time, and their ability to model complex physical phenomenon (Sze et al., 2017). Deep learning models can exploit a particular compositionality in the input features by finding more abstract features in them (Bengio et al., 2021)".

One the other hand, one disadvantage of deep learning models is their lack of explainability i.e. it is difficult to explain their output. In fact, this is an open research problem and many solutions are being proposed to answer this question. We have added this point in section 3.6 (Limitations and future research) of the manuscript.

Lines 427 – 434: "The deep learning-based approach can yield high model performance but it has the limitation in terms of explainability and interpretability (Molnar, 2020). The neural networks are generally considered as black-box and the question of interpreting them is still an open research problem (Mitchell, 2021; Tiddi, 2020). Several methods have been proposed to interpret the behavior of neural networks (Molnar et al., 2020). Explaining the output of neural networks can enhance the confidence of decision makers (Lipton, 2018). Therefore, we propose future research involving deep learning models will benefit if the questions of interpretability and explainability are considered along with model's prediction performance".

Line376-378, There are no Figs. S2a-S2c in Figure S2.

Response: This was mistake to refer Figure S2 here. The actual figure is S1. 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)".

In Fig. S11d and S11e, a peak on 2015-08-28 was captured by both models. However,

It look no data of E. Coli. This should give some explanation if no observed data.

Response: We agree with the reviewer that the peak shown by both HSPF and LSTM in Fig. S11 on August 28, 2015. The observed missing *E. coli* concentration during this peak is likely due to lack of observation. We have mentioned this in the manuscript.

Line 371 – 377: "We observed that both HSPF and LSTM simulated peaks even when the observed data did show corresponding peaks (Figures S8 and S11). The peaks predicted in Fig. S8 are solely from HSPF while the peak event in Fig. S11 is predicted by both HSPF and LSTM models. This shows the efficacy of both calibrated models. We could conclude from Fig. S11 that the lack of observed peak is more likely because of missing observation. However, a similar conclusion cannot be drawn for all the predicted *E. coli* peaks in Fig S8 because of contradicting results of LSTM and HSPF".

It is unclear where input data sources are from for both LSTM and HSPF. Furthermore, is land use resolution same for both LSTM and HSPF?

Response: The input data for both HSPF and LSTM consists of climate, hydrological, E. coli source, and electrical conductivity data. The climate data was measured with an automatic weather station located at the study site. This has been described in following lines in manuscript.

Lines 120 - 122: "Rainfall, relative humidity, solar radiation, wind speed, and air temperature were measured with an automatic weather station Campbell Scientific BWS200, which was equipped with ARG100 (a 0.2 mm capacity tipping bucket)".

The electrical conductivity data was used to calculate surface and sub-surface flow. This method has been used in several previous studies such as Ribolzi et al., 2018. We have given the detailed description of this method in supplementary information (Text S2).

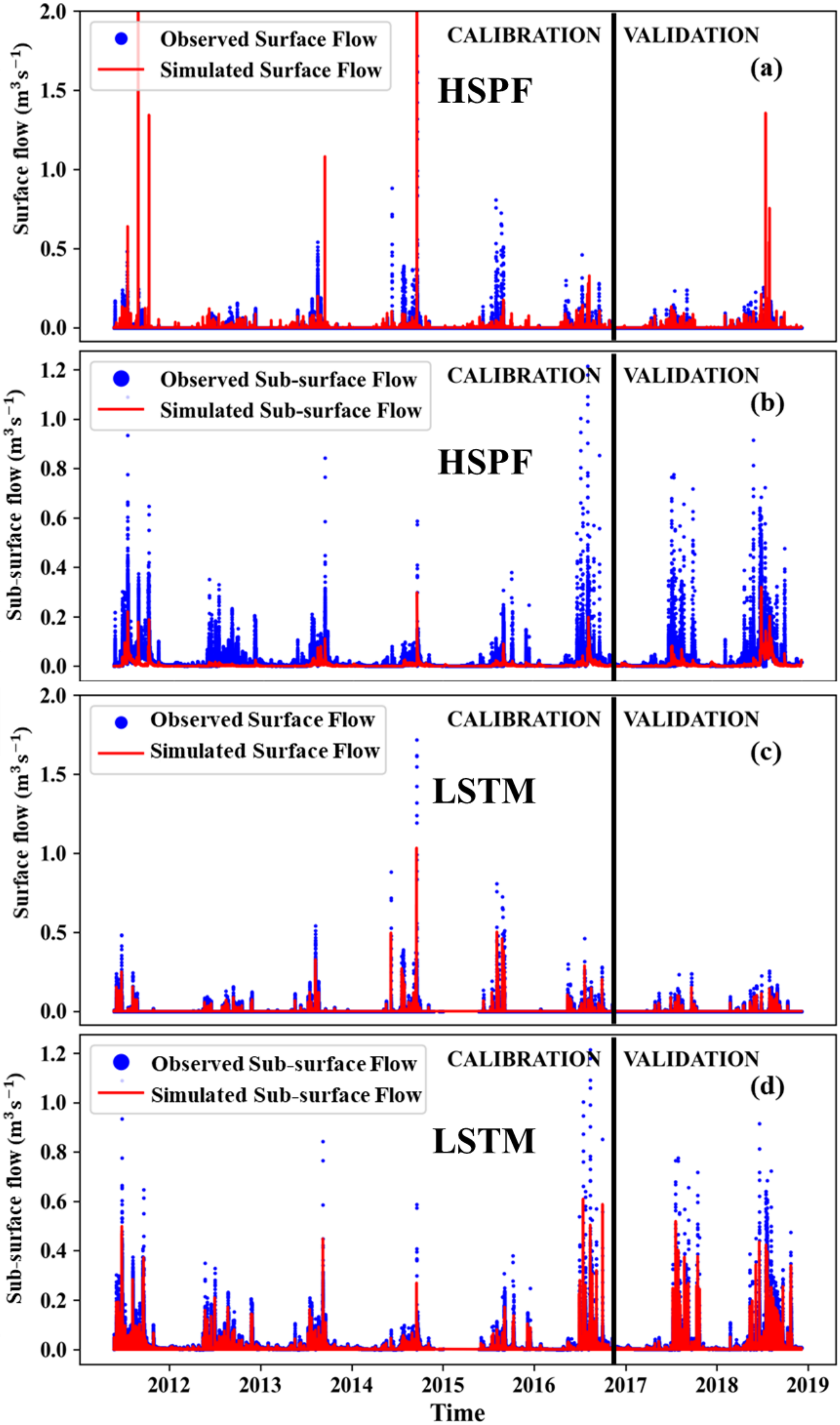
Lines 123 – 128: "We measured the stream water level at the monitoring station using a V-notch and water-level recorder (OTT Thalimedes). The discharge was estimated based on the rating curve relating discharge to water levels. The surface and subsurface flow were calculated using the electrical conductivity method (Ribolzi et al., 2018). A detailed description of this method is provided in the supplementary information (Text S2)".

The *E. coli* concentration was measured based on the standardized microplate method (ISO 9308-3). The detailed description of the experiments is given in supplementary information provided with the manuscript.

Lines 128 – 134: "*E. coli* concentration was measured based on the standardized microplate method (ISO 9308–3). A detailed explanation of *the E. coli* experiment can be found in the supplementary information (Text S3). In this study, we carried out biweekly grab sampling of *E. coli* from 2011 to 2018. Over the same period, we also monitored 11 flood events to assess *E. coli* dynamics during flood events using an automated sampler (ICRISAT) triggered by the water level recorder to collect water after every 2 cm water level change during flood rising and every 5 cm water level change during flood recession".

Fig. 4 and Fig. 5 can be merged to remove a rainfall figure.

Response: We have merged Fig. 5 and Fig. 5 into one Fig. 4. Fig. 4 now shows surface and sub-surface flow from HSPF and LSTM models. The subsequent numbering of all the figures in the manuscript has been updated.



Observed Surface Flow
Simulated Surface Flow
CALIBRATION
VALIDATION
HSPF
(a)
Surface flow (m^3s^{-1})
Observed Sub-surface Flow
Simulated Sub-surface Flow
CALIBRATION
VALIDATION
HSPF
(b)
Sub-surface flow (m^3s^{-1})
Observed Surface Flow
Simulated Surface Flow
CALIBRATION
VALIDATION
LSTM
(c)
Surface flow (m^3s^{-1})
Observed Sub-surface Flow
Simulated Sub-surface Flow
CALIBRATION
VALIDATION
LSTM
(d)
Sub-surface flow (m^3s^{-1})
2012
2013
2014
2015
2016
2017
2018
2019
Time

Figure 4: Hydrological simulation from HSPF and LSTM: (a) Simulated and observed surface flow from HSPF, (b) Simulated and observed sub-surface flow from HSPF, (c) Simulated and observed surface flow from LSTM, and (d) Simulated and observed sub-surface flow from LSTM.

Line 182, it should be briefly described how the data has been converted to a 6 min frequency.

Response: The rainfall data was recorded at 6-minute interval for 2011 and 2012. It was recorded with 1-minute frequency from 2013 to 2018. We used cumulated sum of rainfall data from 2013 to 2018 to convert it into 6-minute time-step. Using the automatic weather station, we also recorded hourly relative humidity, solar radiation wind speed and air temperature. These data were then used to calculate potential evapotranspiration using the method of Penman-Monteith at a 1-h time-step. Finally, we interpolated the potential-calculated evapotranspiration to a 6-min time-step. For E. coli, we considered the values nearest to 6-minute time-step as representative of that time-step. We have briefly discussed this in following lines in the manuscript.

Lines 186-190: "This was carried by interpolating the hourly weather data. Rainfall data were already available at 6 min for 2011 and 2012 while for 2013 to 2018 it was available at 1 min frequency and was aggregated into a 6-min time series. For E. coli concentration, the values nearest to a 6^{-}min step were used as representative of that time step".

Figs. S6-S11 should be explained and discussed.

Response: We have added discussion about hydrological results of HSPF and LSTM which are illustrated in Fig. S6-S11 in the manuscript.

Line 322 – 327: "The simulated surface flow by HSPF followed the rainfall events more closely as compared to that of LSTM. The peaks in surface flow in Fig. S8 are completely missed by LSTM while captured by HSPF model. We also observed that LSTM can follow the observed trends in surface and subsurface flow more closely than the HSPF (Fig. S6, S9, S10). The falling limb from the predicted sub-surface flow of LSTM is gentle and follows the observed pattern (Fig. S9 - S11)".

A discussion paragraph about results of E. coli in Fig. S6 – S11 have been added in response to reviewer's previous comment.

Line 371 – 377: "We observed that both HSPF and LSTM simulated peaks even when the observed data did show corresponding peaks (Figures S8 and S11). The peaks predicted in Fig. S8 are solely from HSPF while the peak event in Fig. S11 is predicted by both HSPF and LSTM models. This shows the efficacy of both calibrated models. We could conclude from Fig. S11 that the lack of observed peak is more likely because of missing observation. However, a similar conclusion can not be drawn for all the predicted *E. coli* peaks in Fig S8 because of contradicting results of LSTM and HSPF".

References

Abimbola, O. P., Mittelstet, A. R., Messer, T. L., Berry, E. D., Bartelt-Hunt, S. L., and Hansen, S. P.: Predicting Escherichia coli loads in cascading dams with machine learning: An integration of hydrometeorology, animal density and grazing pattern, Science of The Total Environment, 722, 137894, 2020.

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

Cho, K. H., Pachepsky, Y. A., Kim, M., Pyo, J., Park, M.-H., Kim, Y. M., Kim, J.-W., and Kim, J. H.: Modeling seasonal variability of fecal coliform in natural surface waters using the modified SWAT, Journal of Hydrology, 535, 377-385, 2016.

Jeon, D. J., Ligaray, M., Kim, M., Kim, G., Lee, G., Pachepsky, Y. A., Cha, D.-H., and Cho, K. H.: Evaluating the influence of climate change on the fate and transport of fecal coliform bacteria using the modified SWAT model, Science of the Total Environment, 658, 753-762, 2019.

Ligaray, M., Baek, S. S., Kwon, H.-O., Choi, S.-D., and Cho, K. H.: Watershed-scale modeling on the fate and transport of polycyclic aromatic hydrocarbons (PAHs), Journal of hazardous materials, 320, 442-457, 2016.

Lipton, Z. C.: The Mythos of Model Interpretability: In machine learning, the concept of interpretability is both important and slippery, Queue, 16, 31-57, 2018.

Meshesha, T. W., Wang, J., and Melaku, N. D.: A modified hydrological model for assessing effect of pH on fate and transport of Escherichia coli in the Athabasca River basin, Journal of Hydrology, 582, 124513, 2020.

Mitchell, M.: Why AI is harder than we think, arXiv preprint arXiv:2104.12871, 2021.

Molnar, C.: Interpretable machine learning, Lulu. com2020.

Molnar, C., Casalicchio, G., and Bischl, B.: Interpretable machine learning–a brief history, state-of-the-art and challenges, Joint European Conference on Machine Learning and Knowledge Discovery in Databases, 417-431, Pachepsky, Y. and Shelton, D.: Escherichia coli and fecal coliforms in freshwater and estuarine sediments, Critical reviews in environmental science and technology, 41, 1067-1110, 2011.

Ribolzi, O., Lacombe, G., Pierret, A., Robain, H., Sounyafong, P., De Rouw, A., Soulileuth, B., Mouche, E., Huon, S., and Silvera, N.: Interacting land use and soil surface dynamics control groundwater outflow in a montane catchment of the lower Mekong basin, Agriculture, Ecosystems & Environment, 268, 90-102, 2018.

Sowah, R. A., Bradshaw, K., Snyder, B., Spidle, D., and Molina, M.: Evaluation of the soil and water assessment tool (SWAT) for simulating E. coli concentrations at the watershed-scale, Science of the Total Environment, 746, 140669, 2020.

Sze, V., Chen, Y.-H., Yang, T.-J., and Emer, J. S.: Efficient processing of deep neural networks: A tutorial and survey, Proceedings of the IEEE, 105, 2295-2329, 2017.

Tiddi, I.: Directions for explainable knowledge-enabled systems, Knowledge Graphs for eXplainable Artificial Intelligence: Foundations, Applications and Challenges, 47, 245, 2020.

Wang, L., Chen, J., and Marathe, M.: Tdefsi: Theory-guided deep learning-based epidemic forecasting with synthetic information, ACM Transactions on Spatial Algorithms and Systems (TSAS), 6, 1-39, 2020.