

Comment on hess-2021-98

Anonymous Referee #1

Referee comment on "In-stream *Escherichia coli* modeling using high-temporal-resolution data with deep learning and process-based models" by Ather Abbas et al., Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2021-98-RC1>, 2021

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 and 0.59 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.

- Line 62, add g in “determining”.
- 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* (Sowah et al., 2020; Meshesha et al., 2020), and limitations of process-based models (Wang et al., 2020, 2021).

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>

- Line85-95, it is unclear what advantages DL has over process-based models. Are there any disadvantages of DL, compared to process-based models?
- Line376-378, There are no Figs. S2a-S2c in Figure S2.
- 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.
- 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?
- Fig. 4 and Fig. 5 can be merged to remove a rainfall figure.
- Line 182, it should be briefly described how the data has been converted to a 6 min frequency.
- Figs. S6-S11 should be explained and discussed.