Comment on hess-2021-98

Anonymous Referee #2

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-RC2, 2021

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.

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. I think that the basin area in this study may be relatively small. What if LSTM model application in a large-scale 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.

Line 29, full name is needed for "PDR"

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

In study site description, basin area is needed.

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

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.

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

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

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

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?

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