

Interactive comment on “HESS Opinions: Deep learning as a promising avenue toward knowledge discovery in water sciences” by Chaopeng Shen et al.

S. Mylevaganam

mylevaganamsivarajah@gmail.com

Received and published: 30 April 2018

- 1) Even though the word “hydrology” is meant for science (“logy”) of water (“hydro”), in the current version of the manuscript, the boundary between water sciences and hydrological/hydrologic sciences is not visible. As per the title of the manuscript, the manuscript is about deep learning in water sciences. However, the content of the manuscript is merely constrained to hydrological sciences (e.g., see P-1 LN-22).
- 2) What is meant by “HESS” opinions? Are the opinions echoed in the manuscript represented by the journal office? Do the authors represent the journal office? Do the authors represent the editorial board? What is the expected outcome of the review

C1

process of this manuscript? If the null hypothesis is that the HESS opinions are always published regardless of the review process, what should be the appropriate alternate hypothesis that needs to be tested for a given significance level? Are the available data sufficient to conduct this hypothesis testing?

3) As per the authors, deep learning, which has gained widespread attention since 2012(see P-2 LN-1), is a suite of tools centering on artificial neural networks. Is there a specific reason for the authors to prefer the year of 2012? In my opinion, the fusion of information theory and deep learning in hydrological sciences was well rooted even before 2012. Therefore, an appropriate reference is needed to support the authors’ statement.

4) The titles of some of the subsections are not acceptable at a significance level of 5%. For example, the title of subsection 2.1 is with more data, opportunities arise. What is the HESS opinion on writing titles for sections/subsections?

5) As per the authors, compared to classical DL problems, hydrology has a unique set of challenges that are research opportunities for DL (See P-10 LN-16). In the subsequent sentence, the authors state that DL research has not cover these questions extensively. What are those questions? I think, the paragraph (P-10 LN-16) needs to be re-written.

6) As per the authors, DL models have already been used as surrogate models for PBMs, but many novel ways that couple the two (i.e., PBMs and DLs) should be investigated (see P-11 LN-18). In my opinion, this has already been investigated in one of the PBMS (SWAT) in hydrology.

7) As per the authors, the evidence is mounting that when given “enough data”, DL can provide the “unique ability” to automatically extract features, sometimes “better than human experts” do(see P-4 LN-28). Subsequent to this statement, the authors provide few bulleted points. What are meant with those bulleted points? Are the bulleted points meant to show that when given “enough data”, DL can provide the “unique ability” to

C2

automatically extract features, sometimes “better than human experts” do?

8) Should the abbreviation ML (P-5 LN-33) be introduced in one of the previous pages (see P-3 LN-6)?

9) On P-3 LN-20, except for satellite-based data products of precipitation, references are given for all other large available datasets (e.g., soil moisture, evapotranspiration, and streamflows) mentioned in the manuscript. Is there a specific reason for not citing a research paper for satellite-based data products of precipitation?

10) With the emerging datasets, DL models can be built and trained to learn features, organizational patterns and relationships and predicts outputs given new input instances (P-3 LN-28). However, the authors are not advocating a whole transition to DL as some of the problems, specifically the problems with just not enough data to train DL-based models, could be best tackled by specifically designed earlier-generation models. I think, it would be more appropriate to show an example (may be in hydrology) of how to use DL models and how to use specifically designed earlier-generation models to avoid transition to DL.

11) Considering the number of authors listed in the manuscript and the quantity of the work carried out in the manuscript, I think, it becomes vital to list each author's contribution in the manuscript.

12) What is meant by citizen “scientists”? What is the minimum required qualification? Does the definition of citizen scientist vary spatially and temporally?

Minor Comments P-3 LN-19: should it be Srinivasan, 2013 P-4 LN-5 to P-4 LN-9: The language needs to be checked

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., <https://doi.org/10.5194/hess-2018-168>, 2018.