AR3 McCabe

Review of Shen et al. (2018) "HESS Opinions: Deep learning as a promising avenue towards knowledge discovery in water sciences" Overview The opinion article by Shen et al. (2018) aims to provide a perspective on the opportunities that deep learning may provide to the water sciences discipline. This is certainly a topic of much current interest and relevance to the community, as it offers possible new pathways for system interpretation and understanding. As such, I was very keen to read and review this contribution, with hopes of 1) learning more about deep learning applications in the "water sciences", and 2) identify some practical outcomes that could be relevant to my own (relatively broad) research interests.

Since I was not previously aware of this type of "Opinions" forum in HESS, I was expecting something more akin to a "Review Article", where the advances in machine learning (being delivered in this case by deep learning) would be illustrated through some relevant applications and examples. As such, I was a little disappointed that this was not the intent of this paper. The paper is precisely as the title dictates: an opinion article. Having, and expressing, an opinion is great: but for it to appear as a published article, it should ideally be supported by a strong, reasoned and defensible position that counters competing arguments via illustration of its superiority (or at least equivalence).

Many thanks to Dr. McCabe who gave very constructive criticism. Indeed this is an opinion paper which does not normally assume the role of a full review. Of course, some concise arguments are provided, but the function of a full review paper has been achieved in another open-access paper on arxiv. Few opinion papers carry a full review and we would like the paper to be shorter. We must take elements (in the form of summaries and abridged examples) from the review paper to support the arguments. In addition, this paper also has the important task fo discussing what the community has to do as a whole

During revision (if allowed), we plan to reorganize the paper in the following fashion:

Section 1. Overview

current overview, with more discussion about promising attributes of DL

Section 2. The emergence of a complementary research avenue

More examples of DL success, some potential uses of DL in hydrology. Some possible interrogative study methods to show the promise.

Section 3. Challenges and opportunities

Expand on original section 4. There are many old and new challenges, many of which cannot be resolved by individual research groups: regionally-imbalanced dataset; strong heterogeneity and contextual variables; partial observations; computational challenges; data access; myriad configurations and "tricks"; lacking training data, especially unlabeled data; problem complexity; missing dynamics; large variation in performance based on DL configurations; Non-stationary world and increasing extremes are beyond previous observations.

Section 4. A community roadmap to DL-powered scientific advances in hydrology. How to solve challenges raised in Section 3 with a community-based approach

(i) synergy between PBM and DL
(ii) readily **accessible large dataset** with uniform formats: earth observations and monitoring networks. assimilate large amount of data to learn true patterns.

(iii) community-shared baseline DL models and data-processing pipelines
(iv) Open and transparent modeling **competitions** in water to facilitate algorithm comparisons, with evaluation on both **performance** and **interpretation** → we need to recognize the significant roles played by competitions in the development of DL research.
(v) Develop a baseline suite of DL interpretation and visualization software that support mainstream DL models, especially those that interpret the hidden layers.

Sections 3 & 4 will be greatly enhanced.

Deep learning may be (and I believe it is) "a promising tool toward knowledge discovery in [the] water sciences". But simply stating it and illustrating with some examples where it has worked before is not the way to convince a new audience
What is presented is a brief description of deep learning, a rather concise historical review of "machine learning" applications in hydrology (e.g. SVM, CART, RFs; which actually have quite an extensive history in hydrology, and especially remote sensing that could be detailed further), an expression of the need for data-driven science in contrast to a more classical (physics-based) approach, and an overview of some of the unique challenges that the water sciences present (which do not seem particularly unique if posed across the earth sciences). However, none of the expressed opinions are particularly revolutionary ideas: hydrology (and related fields) already provides many examples of data-driven science, black-box modeling applications, and novel statistical approaches to divine process insights. What would be good to see is how deep learning transcends these, or at the least, builds upon them to provide an avenue for new insights and investigation into "hydrological" processes.

We do not think interrogative studies was part of the historical studies or even an aspect that was ever argued for in hydrology. These studies were motivated by the criticism that DL models are black boxes, and the computer science community and the general scientific community only recently have worked together to develop these interpretive methods.

DL are non-deep machine learning are substantially different in their capability, data demands and scope. Their differences have been detailed in some DL papers such as Goh et al., 2017, Tao et al., 2016, Fang et al. 2017 andis described at lengths in Shen 2017 (and a revised version of Shen's review paper that is to be posted on arxiv soon). First, the ability to automatically extract or engineer features is a new characteristic of DL that was not available before. The hidden layers provide a new way of providing scientific understanding. Second, the expressive power and the abilities to capture spatio-temporal dependences and high-dimensional data distributions are new possibilities that were not possible before. Transfer learning reduces the demand for data and is more widely exploited with DL.

These characteristics leads to a new plausible way of doing science as we are advocating here. However, it does seem like this has not come out very apparently in the manuscript. We will enhance this part of the discussion with summaries from Shen's review as well as examples not found from there.

Overall, I think there is a missed opportunity here to provide a perspective that could potentially garner significant interest in the community. To do this, the authors could expand on a possible road-map on future directions (and obstacles) for deep learning applications,

and also provide a demonstration of some analogous examples (perhaps from other disciplines, if not from hydrology directly) that could be relevant to "water science" applications. It's my hope that the authors can consider some of my comments in adapting their opinion piece – and ultimately attract the impact such a topic deserves
Comments (in no specific order of importance or logical sequence).

If a revision is permitted we will indeed add a New Section 4 as indicated above, which includes a roadmap of several important pieces. Some of this was covered in the original manuscript, but was not presented in the format of a road-map. Please see our first reply to Dr. McCabe. On the other hand, in this paper we would like to stress the importance of a community-coordinated approach towards solving these obstacles.

**** The title is very broad, with "water sciences" encapsulating a wide range of possible research avenues. I guess this is fine, as I agree that deep learning has broad application, but I wonder whether it might help to focus this discussion on "hydrological" sciences instead, and illustrate with some demonstrations of where this approach might deliver upon its potential. If the title is retained, it would need a much broader description of approaches and applications that could be explored. The authors might wish to review the recent work of Marcais and de Dreuzy (2017), who present a brief introduction to deep learning, focused on some more specific applications (calibration, hypothesis testing, etc.).***

We will indeed change the title to focus on hydrology.

Marcasis and de Dreuzy will be added to the literature review. As described in our reply to reviewer #1, it is a brief "call into the wild".

vv
**** I would remove the repeated statement (see line 24 as an example) of ". . .we lay out several opinions shared by the authors". In fact, I'd remove the use of "opinions" throughout the manuscript (Pg3-L4; Pg3-L9; Pg8-L25; Pg11-L26 etc.) completely and just focus on the presentation of ideas. As an alternative, use instead "Here we propose. . .". However, it is assumed that all co-authors are in agreement with the content of the paper, so there's no need to remind the reader of this.***

Good point. This sentence will be changed.

**** The five points listed in the abstract lean a little towards motherhood statements. Some specificity here would be great. Outlining what "may" happen seems a bit counterproductive. If this is a strongly held "opinion", this should be reflected in the content of the paper. For instance, "Deep learning will revolutionize our understanding of XYZ. . ." or "Deep learning offers an entirely new approach to ABC. . .". At the least, these statements need to be supported throughout the manuscript by a clear and ratio nal review of how (precisely) deep learning will deliver upon them. Point 4 is probably the most important here, and the manuscript could really be built up around this (a point I will discuss below). I do not really understand Point 5 i.e. we need hydrologycustomized methods for interpreting knowledge provided by deep learning? Isn't one of the points of deep learning to provide new knowledge for interpreting hydrological processes? Are you suggesting that it can do this, but we aren't able to understand it? Perhaps it's just me, but I find this a bit confusing.***

We will revised our abstract with more clearly-defined statements.

In terms of the hydrology-customized interpretation method:

According to our summary of Shen 2017's review paper, there are several methods that have already been developed for standard image recognition problems (relevance backpropagation, approximation using interpretable models, and correlation-based analysis). These methods can be ported to water applications. However, scientists in other domains have been creative in devising problem-specific ways in interpreting the results. We believe water scientists need to do both. Especially, considering we will also need customized network structure and our applications will be diverse, some of the method will not work out of the box. Therefore, we expect customized interpretive methods to be necessary and will be an active area of research. Nevertheless, we will re-write this sentence to make it more clear:

(5) An important research need is to customize DL interpretive methods for hydrologic research.

**** Regarding Point 4. To me, this represents the key issue that much of the paper can be built around. Deep learning has potential, but there are some specific challenges that hydrological sciences present that need to overcome or addressed. These are detailed somewhat in Section 4, but so much more could be written and the ideas expanded upon. For deep learning to have an impact in "water science", it is precisely issues like these (and this list is not comprehensive) that need to be considered. It would be great if you could structure your paper to examine these in more detail (if not provide possible solutions or avenues to address them). At the moment, the paper basically says that deep learning is a great technique that has much potential to provide new insights and understanding – BUT – there are some pretty serious roadblocks and challenges (not unique) to hydrological sciences that need to be addressed first. It's a big "but", especially if no attempt to provide a pathway to addressing them is offered. The real value of this opinion piece could be to provide some roadmaps towards these. At the least, a number of the "questions" presented in this section can be examined in greater detail, with examples drawn from the existing literature to showcase earlier or preliminary efforts.***

This is a wonderful suggestion. We believe the manuscript would benefit from such a re-structuring. As indicated by the previous comment, we are in the process of revising the manuscript and we anticipate the structure will look like this:

Section 3. Challenges and opportunities

Section 4. A community roadmap to DL-powered scientific advances in hydrology. How to solve challenges raised in Section 3 with a community-based approach

**** The paper could really use a review of the structure combined with a sharper focus on the deep learning applications (to hydrology/water science) in general. The entire Overview section reads as a Deep Learning review, rather than an exploration of its application to water sciences. Section 2.2. could probably be incorporated into the Overview/Introductory section instead of standing alone. Further, while an in troduction to the technical concept is certainly required (and also needs attention), there's not much in the way of expounding on knowledge discovery. Just as illustrating some examples in other disciplines is relevant and required, so too is exploring those applications already examined in the "water sciences" through some***

recent literature (see your own listed examples on Pg3-L3 as well as on Pg5-L22-28). Providing some brief review of these applications may serve to demonstrate the value of your opinion. There are also quite a few others (see Agana and Homaifar, 2017 10.1109/SECON.2017.7925314)

As summarized in Shen 2017, there really are not a great deal of water applications for DL, and hence this paper. Those have been collected into that review paper. Here we will add a paragraph to summarize the findings, and add some examples. Again please note this paper moves beyond the review: we raise several opinions on how, together as a community, tackle the obstacles. Selling DL is not the only focus of this piece.

**** Following this point, the companion paper of Shen (2017), purports to provide a more comprehensive technical background (it is not listed in the bibliography). I was able to find this on arxiv (https://arxiv.org/ftp/arxiv/papers/1712/1712.02162.pdf) with the title "A trans-disciplinary review of deep learning research for water resources scientists". While only skimming that paper, I can see that it addresses many of my criticisms of this manuscript, in that it provides the needed level of technical background, disciplinary context and demonstration via examples that I was hoping for. The obvious question then is what additional value this manuscript offers in light of that work? I will leave it up to the authors (and editor) to make that assessment [but in the same vein, the EOS article by Shen, 2018, https://doi.org/10.1029/2018EO095649 seems another example of an opinion article on this topic?).***

Please see the table in the reply to AR1 about the differences between these two papers. The review paper provides basics, literature review, and discussion of the hydrologic science challenges that DL could help address. The current Opinions paper focuses on the complementary scientific avenue, the challenges facing application of DL, and what we can do as a community to tackle these challenges. This paper is more forward looking and it is, as titled, an opinion paper. We agree this should be made more clear in the abstract and in the paper. On the other hand, the EOS article, constrained by 600 words limit, was to serve as an opener of a special issue in WRR. It stresses the argument that DL may one day become part of hydrology. There was barely any discussion in that one so it should not really compared.

Table. Difference between papers

Paper	Unique ideas
This HESS Opinion	(As its title indicates, this is truly an opinion paper. We need to assume readers have access to Shen's review paper) 1. Opinion: DL is not a hype. Supported by a review of its solid progress, winnings of competitions and adoption in daily uses. 2. Proposition of the complementary, data-driven scientific avenue: the integration of interrogative studies into the avenue. 3. Following unique Opinions are about what we can do as a community: a. (**to be enhanced**) incorporate PBM and DL b. scientific methods: hypotheses come from machine learning. We do not pose an opinion before doing data mining. Difference from earlier ML: now we have DL to automatically extract features. c. call for open competition of DL in hydrology with criteria focusing on both performance and explainability d. collecting big data through data sharing and citizen scientists 5 (**to be enhanced**). Water science provide unique challenges and opportunities for DL. 6 (**to be added**): Roadmap toward DL-supported science discovery. & Practice challenges and research thrust as a community
Marcasis and Dreuzy 2017	Main points: DL can be used for prediction issues; it may contribute to initial choice and alternatives of physical model structures; model reduction; emergent system properties; calibration. (however, each was only mentioned in one sentence). Test on hydrologic numerical data; benchmarks
Shen. 2018 Review	1. Technical details on ML and DL, regularization, etc. 2. Trans-disciplinary review of DL applications and experiences in various disciplines of sciences 3. Technical details of progress: interpretive DL and GANs 4. (revision) prospects for DL to help tackling grand challenges facing water sciences: inter-disciplinarity, human dynamics, data deluge (from novel sources), scaling and equifinality issues, non-unique inversions and high-dimensional, multi-modal data.

**** Page 2, Line 15-16. This sentence is unclear to me***
**** Some of the short-comings of GANS should also be mentioned: especially their ability to be***

"easily fooled" (see https://arxiv.org/pdf/1801.00553.pdf, https://arxiv.org/abs/1710.09762 and many other similar papers). Are there implications to water sciences in this – especially for automated approaches used in prediction systems? What other drawbacks of deep learning may impair their uptake and development?
**** Other papers that might be of interest to the authors (indeed, see Volume 55, Issue 5 of Groundwater):***

We will add some discussion here. It is not that GANs are easily fooled, but deep networks can be fooled by small changes in inputs. GANs can actually be used to train networks more robustly, which is one of the co-authors focus of research [Ororbia].

Chen and Wang (2018) "Recent advance in earth observation big data for hydrology" https://doi.org/10.1080/20964471.2018.1435072 Frere (2017) "Revisiting the Relationship Between Data, Models, and DecisionâĂ˘ R- ˘ Making" https://doi.org/10.1111/gwat.12574 Lary et al. (2016) "Machine learning in geosciences and remote sensing" https://doi.org/10.1016/j.gsf.2015.07.003 Marshall (2017) "Creativity, Uncertainty, and Automated Model Building", https://doi.org/10.1111/gwat.12552 Lidard et al. (2017) "Scaling, similarity, and the fourth paradigm for hydrology", https://doi.org/10.5194/hess-21-3701-2017 Anderson (2008) "The end of theory: the data deluge makes the scientific method obsolete" https://www.wired. com/2008/06/pb-theory/ McCabe et al. (2017) "The future of Earth observation in hydrology", https://doi.org/10.5194/hess-21-3879-2017

**** Since I'm familiar with that last reference, I highlight some of the discussion therein on machine learning approaches in general, particularly on Page 3902 (n.b. it may also be worth reviewing some of the mentioned references in an attempt to provide context of machine learning based hydrological applications - and where deep learning will fit into that): "Despite this remarkable confluence of data science and remote sensing, one can still resist the narrative that there is no problem that a sufficiently complex machine-learning algorithm cannot unravel given enough data (Anderson, 2008). If this were the case, there would be no need for domain expertise to understand current and future challenges in hydrology: the dilettante will have prevailed (Klemeš, 1986). Indeed, there remain several obstacles to any predicted ascension of a completely data-driven approach to hydrology. Observations of the hydrosphere often have a spatio-temporal structure that emerges in the form of correlations between variables, but this correlation may not necessarily imply causality. Therefore, being able to draw strong deterministic conclusions about the behaviour of hydrologic systems based on data-driven methods often requires prior knowledge (and understanding) of the physical processes (Faghmous and Kumar, 2014)." This is relevant to your Section 2.4 and elsewhere***

Some of these points indeed resonate with what we have put forth in old Section 4 where we mentioned PBMs and DL models complement each other. We in fact do not argue for a completely data-driven scenario where domain expertise is of no use. This is not from a personal interest point of view, but we do not see water science presents so much data that can cover every aspect of the hydrologic and human water cycle. If we humor ourselves and imagine that such a scenario does occur, it may indeed be possible for DL models to predict everything more accurately than process-based models, but still PBMs are required for us, humans, to understand the causal relationships. Even data in the world may be not be sufficient to distinguish between causal and associative relationships.

In our revision we will make these points more clear, and will cite the above-mentioned literature so our viewpoints can be put into context.

**** Your Section 4 provides an excellent launching point to really expand on some of these ideas and challenges (see above), and I would encourage you to use these (and build upon them) to structure this opinion piece around. Of course, it should be recognized that the problems highlighted here are not particular to deep learning, but to hydrological inference and understanding broadly, and that there has been much effort directed towards novel statistical approaches to address some of these (which would be worth mentioning, or at least providing some context)***

Thanks and should a revision be allowed, we will expand this section into a much bigger section. We will update this online response as revisions are done.

**** I'm not convinced that Section 2.4 is essential to this paper – or at least it can be presented differently. Advocating the role of data-driven approaches is not a new concept in hydrology (see some of the papers above for reviews) – nor is it especially controversial. It is not like modelers act in isolation – data is an integral part of that process. As with the use of machine learning approaches, data-driven knowledge discovery has a rich history in hydrology, which may be worth reviewing. Certainly there are many examples of ANN type models outperforming their physically-based counterparts. But I'm not sure what the intent of this section is? Either way, it is also not immediately clear (or demonstrated) that deep learning offers a better path towards achieving this "goal" than the myriad of techniques already being used.***

As mentioned previously, we do not believe interrogative studies were mentioned before, as they are essentially a new sub-discipline. Therefore, while we will take most of the suggestions offered by the reviewer, here we beg to differ that this section is important. Indeed, we will outline the novelty more clearly.

**** Likewise, I'm not sure what the purpose of Section 3.2 is? The last paragraph in particular (Pg10-L6-15) invokes a lot of hand-waving***

This is actually one of the roadmap toward collecting more data. Citizen scientists can help collect data about precipitation, groundwater depths and pressure, surface water stages, soil texture and other observable variables. Previously it was not possible to consume lot of the sub-research-grade data they produce. Now, with lots of data, the noise inherent in these data can be averaged out. With DL, we can infer concepts not possible before.

There are a number of other questions I have and handwritten annotations I have made on the paper that are not included in this review. My overall impression is that the paper needs some considered thought not just on its structure, but on how it attempts to present the "opinion" that deep learning is a promising tool in hydrology. While I'm an advocate of your perspective here, in reading the manuscript, I found little to convince me that this approach presents a radical new angle to anything that has come before it. I hope that the authors can

address some of these comments and further refine the contribution, as I think it is a topic that will be of considerable interest to the community.

Again we thank Dr. McCabe for constructive criticism. I think the main focus may be a little different from what Dr. McCabe had anticipated. Here not only do we want to argue the usefulness of DL (partially accomplished elsewhere), we want to address how the community as a whole can incubate DL research.