

Dear reviewers and editors,

We have completed a thorough revision of the manuscript. Major section re-organization have taken place. New Figures and new sections have been added to address the issues. We believe the paper has improved significantly from the first version. In the following, line numbers in the curly brackets refers to the starting position of the excerpts in the Track Change document.

***ARI***

***General comments This article discusses the potential benefit of deep learning models to let emerge knowledge about water science systems from hydrological data. The paper is well written, the opinion is clearly stated and the authors present their arguments based on their expertise and their understanding of deep learning techniques. I'm wondering to what extent this is new and original compared to the opinion paper of Marçais & Dreuzy (2017 - see reference below). For example, the figure presented in this former article expressly conveys the idea that DL methods could enhance the unraveling of hydrological properties from data which is the core of this current article.***

***Marçais, J., & de Dreuzy, J. R. (2017). Prospective interest of deep learning for hydrological inference. Groundwater, 55(5), 688-692.***

The roles taken by these papers are very different. We are cheerful to see others echoing the same enthusiasm for DL in hydrology. We would welcome others to have discussion and work together on this topic. In terms of the paper, though, we see significant differences between these articles, which is summarized in the table below. In our opinion, the Marcasis and de Dreuzy 2017 (MD17) paper was a timely and welcomed first “call into the wild” (with the need to mention that one of the co-authors of this Opinion paper has had a DL paper, Tao et al. in 2016). While it is a great addition, that paper was very brief and had a different focus. It did not explain why DL could unravel hydrologic properties. It did not mention interrogative method which is a crucial part of our argument. It also did not discuss what we need to do as a community to incubate such research. In this article we gather from our past working to voice some enthusiasm as well as challenges.

However, prompted by the reviewer (and other comments), we will revise this Opinion paper significantly to emphasize our main points, which are: (1) DL+interrogative study is a valuable research avenue; (2) what challenges face the community and what we can do together to incubate DL research; (3) water resources present unique challenges and opportunities for DL.

Table. Difference between papers

<p>Table. Difference between papers Paper</p>	<p>Unique ideas</p>
<p>This HESS Opinion</p>	<p>(As its title indicates, this is truly an opinion paper. We need to assume readers have access to Shen’s review paper)</p> <ol style="list-style-type: none"> <li>1. Opinion: DL is not a hype. Supported by a review of its solid progress, winnings of competitions and adoption in daily uses.</li> <li>2. Proposition of the complementary, data-driven scientific avenue: the integration of interrogative studies into the avenue.</li> <li>3. Following unique Opinions are about what we can do as a community:             <ol style="list-style-type: none"> <li>a. 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.</li> <li>b. call for open competition of DL in hydrology with criteria focusing on both performance and explainability</li> <li>c. collecting big data through data sharing and citizen scientists</li> </ol> </li> <li>6. Water science provide unique challenges and opportunities for DL.</li> <li>7. Roadmap toward DL-supported science discovery</li> <li>8. Integration model and DL</li> <li>9. Education</li> </ol>
<p>Marcasis and Dreuzy 2017</p>	<p>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</p>
<p>Shen. 2018 Review</p>	<ol style="list-style-type: none"> <li>1. Technical details on ML and DL</li> <li>2. Trans-disciplinary review of DL applications and experiences in sciences</li> <li>3. Technical details of progress: interpreting DL and GANs</li> <li>4. 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, inverse emergence</li> </ol>

***I also feel that DL techniques and especially why it does work so well is still not understood by computer scientists and mathematicians. However, this article can give the impression that the “DL reasons of success” are now understood (see specific comments) paving the way for knowledge discovery in water sciences through its use. I would consider being more cautious about that as the understanding of the specific properties of DL models compared to more traditional statistical learning models is still an active area of research. This does not mean that DL has not to be widely tested for hydrologic purposes.***

There have been some studies that looked at why DL is powerful, but the point is taken. We added the following paragraph

*“The fundamental theories on why DL generalizes so well have not been maturely developed (Section 2.7 in Shen18). In the ongoing debates, some argued that a large part of DL’s power comes from memorization while others countered that DL prioritizes learning simple patterns [Arpit et al., 2017; Krueger et al., 2017] and a two-stage procedure (training and testing) also helped [Kawaguchi et al., 2017]. Despite these explanations, it has been found in vision DL that deep networks can be fooled by adversarial examples where small, unperceivable perturbations to input images sometimes cause large changes in predictions, leading to incorrect outcomes [Szegedy et al., 2013; Goodfellow et al., 2015]. It remains to be seen whether such adversarial examples exist for hydrologic DL applications.” {p8L8}*

***Specific comments Page 1 L.20 Could you specify articles where DL shows capacities for scientific discovery?***

This is based on example and reasoning. It was included in the review paper Shen2018. There were quite a few examples. Here, we will include some summary of these in the revised Opinion article:

*“Given that deep networks can identify features without human guide, it follows that they may extract features that the algorithm designers were unaware of, or did not intentionally encode the network to do. If we could believe that there is latent knowledge about the hydrologic system that humans are not yet aware of, but can be determined from data, the automatic extraction of features leads to a potential pathway toward knowledge discovery. For example, deep networks recently showed that grid-like neuron response structures automatically emerge at intermediate network layers for a network trained to imitate how mammals perform navigation, providing strong support to a Nobel-winning neuroscience theory about the functioning of these structures (Banino et al., 2018).” {p7L10}*

***Page 2 L.9-16. The paragraph gives the impression that DL is a “plug and play” model whereas to my knowledge building a DL model still requires intensive computer scientists’ knowledges and requires use of GPUs.***

Good point. We added the following sentence to this sentence:

*“While showing many advantages, DL models will require substantial amount of computing expertise. The tuning of hyper-parameters, e.g. network size, learning rate, batch size, etc., often require a priori experiences and trial and error. The computational paradigm, e.g., computing on graphical processing units, is also substantially different from ordinary hydrologists’ educational background.” {P8L3}*

*Page 2 L.24. I don't think that generalization capacities of DL come from its interpolation capability. Indeed, classical neural networks have been proven (see citation below) to be universal interpolators but they do not generalize well. Hornik, K., Stinchcombe, M., & White, H. (1989). Multilayer feedforward networks are universal approximators. Neural networks, 2(5), 359-366. Page 3 L.25. I agree that increase in environmental data opens new opportunities for data-driven techniques in general and particularly for DL techniques. Along with the development of spatialized, remote sensing data, I would also insist on the development of environmental observatories that collect a lot of time series, monitoring data even though they are site specific. These two types of data are complementary to advance through knowledge discovery in hydrology.*

The original sentence was “Moreover, the differentiable nature allows for greater success for interpolation and mild extrapolation, contributing to the strong generalization capability of DL.”. The differentiability “contributes” to the generalization ability, but that is not the sole reason. Other factors include improved architecture, regularization, big data, weights sharing etc., which were mentioned earlier. To avoid any confusion, this sentence has been revised as

“Compared to earlier models like classification trees, most of the deep networks are differentiable, meaning that we can calculate derivatives of outputs with respect to inputs or the parameters in the network. This feature enables highly efficient training algorithms that exploit these derivatives.”  
{P6L24}

*Page 6. L.6-17. This paragraph is intended to bridge the link between interrogative techniques brought in DL by the “AI neuroscience” subdiscipline and the potential of DL for knowledge discovery in water sciences. If the arguments tend to prove that such interrogative techniques enlighten the way the architecture of DL works, it does not explain the success of DL in itself. For example, the sentence*

*L.13: “activations of recurrent neural networks can be visualized to show the control domain of certain cells, which explains its functioning” is not correct. This only explains the functioning of the architecture of the DL, not the reason of success of such a method. There is some literature exploring the need for explanation of DL techniques. For Convolutional Neural Networks (CNNs), their understanding can be linked with wavelet theory (see reference below). Especially their capacity to extract invariants through a lot of different scale in high dimensional datasets but this is still a subject of active research. This capacity could explain their generalization capabilities especially for image datasets.*

The interpretive study does not solely focus on “why DL was successful”, and this is not really the point. The interpretive studies answer “what has DL learned”.

The revision has separated the two questions. At the end of Section 2.2, we mention present theories about “why DL generalizes so well”. The whole 2.3 explains “how to extract knowledge from deep networks”. Thus there should not be confusion now.

*Mallat, S. (2016). Understanding deep convolutional networks. Phil. Trans. R. Soc. A, 374(2065), 20150203. Page 7 L.8-18. It could be interesting to explore how DL techniques can improve hypothesis testing through an exploration of competing process-based models? The*

*Structure for Unifying Multiple Modeling Alternatives (SUMMA) (see reference below) could be a start to generate process-based models with alternative hypotheses. For example, process-based models could be used to feed DL models with numerical generated data.*

*Clark, M. P., Nijssen, B., Lundquist, J. D., Kavetski, D., Rupp, D. E., Woods, R. A., ... & Arnold, J. R. (2015). A unified approach for process-based hydrologic modeling: 1. Modeling concept. *Water Resources Research*, 51(4), 2498-2514*

The reviewer is spot-on that the SUMMA framework is relevant and could be integrated with DL models in several ways. In fact it has been cited in Shen18 in a more relevant reference to equifinality.

Here, the new Section 4.1 mentions this

*“Second, PBMs can augment input data for DL models. PBMs can be used to increase supervising data for DLs, for example, for climate or land-use scenarios that have not existed presently, to augment existing data. Given model structural uncertainty (uncertainty with hydrologic processes), frameworks like the Structure for Unifying Multiple Modeling Alternatives (Clark et al., 2015) could be employed to generate a range of inputs.” {P16L23}*

***Page 10 L.20-Page 11 L.24. I would add to this list the fact that water sciences provide to DL a unique challenge because hydrologic data are intrinsically heterogeneous. Building a model able to integrate these heterogeneous data might be the key toward knowledge discovery in water sciences and toward big progresses in AI.***

Good point! This part has been explained to include this point:

*“Hydrologic data are accompanied by a large amount of strongly heterogeneous [Blöschl, 2006] “contextual variables” such as land use, climate, geology, and soil properties. The proper scale at which to represent heterogeneity in natural systems is a vexing problem [Archfield et al., 2015], as micro-scale of soil heterogeneity, for example, is not computationally realistic in hydrologic models. The scale at which heterogeneity should be represented varies with setting and elements of the water cycle [Ajami et al., 2016]. Moreover, while we recognize that heterogeneity exists in contextual features, many of these features, such as soil properties and hydrogeology, are poorly characterized across landscapes, but both features play important role in controlling water movement. Heterogeneity needs to be adequately represented without radically bloating the parameter space of the models. Furthermore, the heterogeneous physiographic factors covary [Troch et al., 2013] and exert complicated causal and non-causal connections, but we have limited knowledge of their covariation. Consequently, training with insufficient data may result in many alternative DL models that cannot be rejected.” {P13L28}*

### ***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 have now been provided, but the function of a full review paper has been achieved in another open-access paper on arxiv.

We have taken elements (in the form of summaries and abridged examples) from the review paper Shen18 to support the arguments. In addition, this paper also has the important task of discussing what the community has to do as a whole

*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.*

With respect to convincing people of the power of DL, the re-organized manuscript added the following:

*“The progress in AI brought forth by DL to various industries and scientific disciplines is revolutionary (Section 4 in Shen18) and can no longer be ignored by the hydrology community. Major technology firms have rapidly adopted and commercialized DL-powered AI (Evans et al., 2018). For example, Google has re-oriented its research priority from “mobile-first” to “AI-first” (Dignan, 2018). The benefits of these*

*industrial investments can now be felt by ordinary users of their services such as machine translation and digital assistants who can engage in conversations sounding like a human (Leviathan and Matias, 2018). Moreover, AI patents of industries and scientific disciplines grew at a 34% compound annual growth rate between 2013 and 2017, apparently after DL's breakthroughs in 2012 (Columbus, 2018). Also reported in (Columbus, 2018), more than 65% of data professionals responded to a survey indicating AI as their company's most significant data initiative for next year. ” {P4L11}*

We believe Table 1 is also quite convincing:

*“Since 2011, the number of entries with DL as a topic increased almost exponentially, showing around 100% compound annual growth rate before 2017 (Table 1). DL evolved from occupying less than 1% of machine learning (ML) entries in computer science (CS) in 2011 to 46% in 2017. This change showcases massive conversion from traditional machine learning to DL within computer science. Other disciplines lagged slightly behind, but also experienced exponential increase. They also saw the DL/ML ratio jumping from 0% in 2011 to 33% in 2017. As reviewed in Shen18, DL has enhanced the statistical power of data in high energy physics, and the use of DL can be considered to be equivalent to a 25% increase in the experimental dataset (Baldi et al., 2015). In biology, DL has been used to predict potential pathological implications from genetic sequences (Angermueller et al., 2016). DL models in computational fed with raw-level data have been shown to outperform those using expert-defined features when they predict high-level outcomes, e.g., toxicity, from molecular compositions (Goh et al., 2017). Just like other methods, DL may eventually be replaced by newer ones, but that is not a reason to hold out on possible progress.” {P4L20}*

Then, in technical advances, we described how DL transcends traditional DL

*“Underpinning the powerful performance of DL are its technical advances. The deep architectures have several distinctive advantages: (1) deep networks are designed with the capacity to represent extremely complex functions. (2) After training, the intermediate layers can perform modular functions which can be migrated to other tasks, in a process called transfer learning, and extend the value of the training data. (3) The hidden layer structures have been designed to automatically extract features, which helps dramatically reducing labour, expertise and the trial and error time needed for feature engineering. (4) Compared to earlier models like classification trees, most of the deep networks are differentiable, meaning that we can calculate derivatives of outputs with respect to inputs or the parameters in the network. This feature enables highly efficient training algorithms that exploit these derivatives. Moreover, the differentiability of neural networks enables querying DL models for sensitivity analysis of outputs to input parameters, a task of key importance in hydrology.*

*Metaphorically, the intermediate (or hidden) layers in DL algorithms can be understood as placeholders for tools that are to be built by deep networks themselves. These hidden layers are trained to calculate certain features from the data, which are then used to predict the dependent variables. For example, Yosinski (2015) showed that some intermediate layers of a deep vision recognition network are responsible for identifying the location of human or animal faces; Karpathy et al., (2015) showed that some hidden cells in a text prediction network act as length counters of a line while some others keep track of whether the text is in quotes or not. These functionalities were not bestowed by the network designers, but emerged by themselves after network training. Earlier network architectures either did not*

have the needed depth, or were not designed in an artful way such that the intermediate layers could be effectively trained. For more technical details, refer to an introduction in Schmidhuber (2015) and Shen (2018).” {P6L20}

These characteristics leads to a new plausible way of doing science as we are advocating here.

“Given that deep networks can identify features without guide, it follows that they may extract features that the algorithm designers were unaware of, or did not intentionally encode the network to do, leading to a potential pathway toward knowledge discovery. For example, deep networks recently showed that grid-like neuron response structures automatically emerge at intermediate network layers for a network trained to imitate how mammals perform navigation, providing strong support to a Nobel-winning neuroscience theory about the functioning of these structures (Banino et al., 2018)” {P7L6}

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.

***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).***

We think that the reviewer’s suggestion is excellent and have adopted the suggestion. The major re-organization in the revision should have been a significant improvement. The new Section 4 and a new Figure expands on the roadmap.

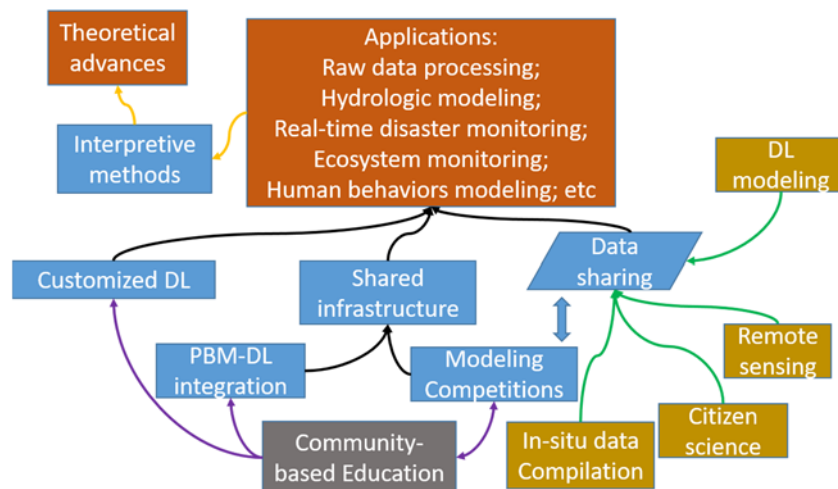


Figure 4. A roadmap toward DL-powered scientific discovery in hydrologic science. Data availability can be increased by (green arrows) collecting and compiling existing data, incorporate novel data sources such as those collected by citizen scientists, remote sensing and modelled dataset. DL can be employed to predict data that are currently difficult to observe. The modelling competitions and the integration between PBM and DL will build important shared computing and analytic infrastructure, which, together



with data sources, support a wide range of hydrologic applications. Interpretive methods should be attempted to extract knowledge from trained deep networks (orange arrows). Underpinning these activities is the enhanced, community-based educational program for machine learning in hydrology (purple arrows). However, these activities, especially the modelling competitions, might in turn feedback to the educational activity.

*\* 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 Marcasis and de Dreuzy (2017), who present a brief introduction to deep learning, focused on some more specific applications (calibration, hypothesis testing, etc.).*

The title has been changed to “HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community”

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”.

*\* 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.*

This sentence has been removed.

*\* 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 rational 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, this sentence has been removed due to total re-writing of the abstract.

*\* 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 restructuring. As indicated by the previous comment, we are in the process of revising the manuscript and the revised structure is:

### **3. Challenges and opportunities**

#### **4. A community roadmap toward DL-powered scientific advances in hydrology**

- 4.1. Integrating physical knowledge, process-based models, and DL models
- 4.2. Multi-faced, community-coordinated hydrologic modelling competitions
- 4.3. Community-shared resources and broader involvement
- 4.4. Education

*\* 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 introduction 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 have added a paragraph as a summary:

*“In contrast, only a handful of applications of big data DL could be found in hydrology, but they already demonstrated great promise. Vision DL has been employed to retrieve precipitation from satellite images, where exhibited a materially-superior performance than earlier-generation neural networks (Tao et al., 2017, 2018). GAN was used to imitate and generate scanning images of geologic media (Laloy et al., 2018), where the authors showed realistic replication of training image patterns. Time-series deep learning network was employed to temporally extend satellite-sensed soil moisture observations (Fang et al., 2017) and was found to be more reliable than simpler methods. Regionalized time series DL rainfall-runoff models have been created (Kratzert et al., 2018). There are also DL studies, based on smaller dataset, to help predict water flows in the urban environment (Assem et al., 2017) and water infrastructure (Zhang et al., 2018). In addition to utilizing big data, DL was able to create valuable, big datasets that could not have been otherwise possible. For example, utilizing DL, researchers were able to generate new datasets for Tropical Cyclones, Atmospheric Rivers and Weather Fronts (Liu et al., 2016; Matsuoka et al., 2017) by tracking them. Machine learning has also been harnessed to tackled the convection parameterization issue in climate modelling (Gentine et al., 2018).” {P6L6}*

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.

*\* 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?*

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].

We added the following

*“Despite these explanations, it has been found in vision DL that deep networks can be fooled by adversarial examples where small, unperceivable perturbations to input images sometimes cause large changes in predictions, leading to incorrect outcomes (Goodfellow et al., 2015; Szegedy et al., 2013). It remains to be seen whether such adversarial examples exist for hydrologic DL applications. If we can recreate adversarial examples, they can be added into the training dataset to improve the robustness of the model (Ororbia et al., 2016).” {P8L9}*

*\* Other papers that might be of interest to the authors (indeed, see Volume 55, Issue 5 of Groundwater):*

*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 Decision 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 argue in the revised paper that DL alone is not going to “do it” in hydrology, in Section 3. 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.

Below are our revised statements. We have included citations suggested by the reviewer.

*“Observations in hydrology and water science generally are regionally and temporally imbalanced.”* {P13L8}

*“Global change is altering the hydrologic and related cycles, and hydrologists must now make predictions in anticipation of changes, beyond previously observed ranges”* {P13L18}

*“Observations of the water cycle tend to focus on one aspect of the water cycle, and seldom offer a complete description.”* {P13L23}

*“Furthermore, the heterogeneous physiographic factors co-evolve and covary (Troch et al., 2013) with complicated causal and non-causal connections (Faghmous and Kumar, 2014). The relationships of soil, terrain and vegetation are further conditioned on geologic and climate history and often do not transfer to other regions (Thompson et al., 2006). Consequently, training with insufficient data may result in overfitted data-driven models or many alternative DL models that cannot be rejected. On the flip side, such complexity due to co-evolution also nearly precludes a reductionist approach where all or most of these relationships are clearly described from fundamental laws”* {P14L6}

As a solution to these issues, we actually promote integrating process-based models and the data-driven one, as described in the New Section 4.1:

#### ***“4.1. Integrating physical knowledge, process-based models, and DL models***

*To address the challenge of data limitations (data quantity), we envision that a critical and necessary step is to more organically integrate hydrologic knowledge, process-based models, and DL. Process-based models, as they are derived from underlying physics, require less data for calibration than data-driven models. They can provide estimates for spatial and temporal data gaps and unobserved hydrologic processes. Well-constructed PBMs should also be able to represent temporal changes and trends. However, because data-driven models directly target observations, these models may have better performance in locations and periods where data are available. Also, as discussed earlier, data-driven models are less prone to a priori model structural error than are PBMs. We should aim to maximally utilize the best features of each type of models.*

....

”

***\* 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. We have expanded this section into two much bigger sections, the new Section 3 and 4.

*\* 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. We believe the new section 2.4 is .

*\* 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. Hopefully the revised version has more clarity and usefulness.

Reviewer #3. Dr. Sawicz

*The manuscript titled “Deep learning as a promising avenue toward knowledge discover in water sciences” conveys the opinion that hydrology is a field well suited for deep learning and that deep learning techniques should be widely applied to increase our understanding of hydrologic systems. While I agree with the authors’ general premise and believe the article can have a great impact on the direction of the technical analysis of future hydrological studies, I also was disappointed to not see two primary topics. The first of which is a summary of how deep learning has been integral in other fields to increase knowledge discovery of other fields of science. The second and more important topic would be the presentation of a general framework of how to apply the techniques of deep learning to open or poorly understood problems within the hydrologic sciences. Section 4 does present some ideas of areas that deep learning may be applied, but a framework of how to apply DL techniques to these problems would help convince the reader of their utility*

*The manuscript is generally well written, but I would recommend more attention to simplifying the verbiage and clarify the message of the paper. To help communicate this further, I included examples within the specific comments below as a guide. With thoughtful revision, I believe that this paper can serve the community well to help show the utility of deep learning techniques. I also think that the inclusion of a companion paper to explain the more technical aspects of deep learning was a very good decision*

Agreed. We have made effort to simplify the paper in terms of terminology. We shall use simpler terms or make some clear definitions. Examples are given below as response to detailed comments.

*Specific Comments: I have included some specific comments that should be revised and used as examples to help guide the authors in their overall revision. Page 2 Line 19: “deep networks are differentiable from outputs to inputs, giving them practical advantages in efficient parameter optimization via backpropagation (training).” It is not clear to me what is meant by differentiable from outputs to inputs. I believe that the concept the authors are trying to communicate here is simple, but it is not done so effectively*

*Page 2 Line 23-24: “Moreover, the differentiable nature allows for greater success for interpolation and mild extrapolation, contributing to the strong generalization capability of DL.” This sentence is very thick in jargon. I would suggest simplifying the verbiage to improve readability and connection to the rest of the paper*

*“Compared to earlier models like classification trees, most of the deep networks are differentiable, meaning that we can calculate derivatives of outputs with respect to inputs or the parameters in the network.” {P6L24} We believe this sentence works better.*

For another example, the main technical parts are now simplified:

*“Underpinning the powerful performance of DL are its technical advances. The deep architectures have several distinctive advantages: (1) deep networks are designed with the capacity to represent extremely complex functions. (2) After training, the intermediate layers can perform modular functions which can be migrated to other tasks, in a process called transfer learning, and extend the value of the training*

data. (3) The hidden layer structures have been designed to automatically extract features, which helps dramatically reducing labour, expertise and the trial and error time needed for feature engineering. (4) Compared to earlier models like classification trees, most of the deep networks are differentiable, meaning that we can calculate derivatives of outputs with respect to inputs or the parameters in the network. This feature enables highly efficient training algorithms that exploit these derivatives. Moreover, the differentiability of neural networks enables querying DL models for sensitivity analysis of outputs to input parameters, a task of key importance in hydrology.” {P6L20}

“Some recent progress in DL research focused on addressing these concerns. Notably, a new sub-discipline, known as “AI neuroscience” has produced useful interrogative techniques to help scientists interpret the DL model (see literature in Section 3.2 in Shen18). The main classes of interpretive methods include (i) reverse-engineer the hidden layers: attributing deep network decisions to input features or a subset of inputs; (ii) transferring knowledge from deep networks to interpretable, reduced-order models. (iii) visualization of network activations. Many scientists have also devised case-by-case adhoc methods, e.g., to investigate the correlation between inputs and cell activations (Shen, 2018; Voosen, 2017).” {P8L25}

***Page 4 Line 9-10: “As a result, over time, some may have grown dispassionate about progress in machine learning, and some may have concerns about whether DL is a real progress or just a “hype.” While I believe that the authors do reflect the sentiment of some within the community to the promise of machine learning, the opinion paper does not present much to dispel these feelings either. In accordance with the mention of my first point in General Comments, it would serve the paper to include some proof as to the utility of machine learning and deep learning***

***Page 4 Line 14-15: “The progress brought forth by DL to the information technology industry is revolutionary (Section 4 in Shen17) and can no longer be ignored.” While a companion paper should compliment this paper, this opinion manuscript should also provide evidence to the point. This could even include a small summary of findings. It is natural to have some overlap between the papers, and I believe that this suggested overlap has purpose.***

The revised Section 2.2 largely fills this role. This section begins with industry uptake, then goes into uptake of DL in computer science and non-computer science research and then list computer science competitions which are now dominated by DL.

We think that some statistics may be convincing in terms of how useful DL is. “Major technology firms have rapidly adopted and commercialized DL-powered AI (Evans et al., 2018). For example, Google has re-oriented its research priority from “mobile-first” to “AI-first” (Dignan, 2018). The benefits of these industrial investments can now be felt by ordinary users of their services such as machine translation and digital assistants who can engage in conversations sounding like a human (Leviathan and Matias, 2018). Moreover, AI patents of industries and scientific disciplines grew at a 34% compound annual growth rate between 2013 and 2017, apparently after DL’s breakthroughs in 2012 (Columbus, 2018). Also reported in (Columbus, 2018), more than 65% of data professionals responded to a survey indicating AI as their company’s most significant data initiative for next year.” {P4L13}



*“The fast growth is clearly witnessed from literature searches. Since 2011, the number of entries with DL as a topic increased almost exponentially, showing around 100% compound annual growth rate before 2017 (Table 1). DL evolved from occupying less than 1% of machine learning (ML) entries in computer science (CS) in 2011 to 46% in 2017. This change showcases massive conversion from traditional machine learning to DL within computer science. Other disciplines lagged slightly behind, but also experienced exponential increase. They also saw the DL/ML ratio jumping from 0% in 2011 to 33% in 2017.” {P4L20}*

Then we gave some concrete examples, but more examples are given in Shen18.

*“As reviewed in Shen18, DL has enhanced the statistical power of data in high energy physics, and the use of DL can be considered to be equivalent to a 25% increase in the experimental dataset (Baldi et al., 2015). In biology, DL has been used to predict potential pathological implications from genetic sequences (Angermueller et al., 2016). DL models in computational fed with raw-level data have been shown to outperform those using expert-defined features when they predict high-level outcomes, e.g., toxicity, from molecular compositions (Goh et al., 2017).” {P4L15}*

***Page 10 Line 20: “Observations in hydrology and water sciences. . .” Some would consider hydrology to be a subset of water science and others may say that hydrology and water sciences are the same field named differently. While cleaning up the language used in the manuscript, I would also suggest using either hydrology or water science. This may be a small point but is one that should be echoed through the paper.***

Yes. now we stick to “hydrology”.

***In addition to these specific comments, I would also encourage the authors to include the various references listed by reviewers 1 and 2. I do not personally have anything to add to these references, but they would serve to present a fuller picture of machine learning applications within hydrology.***

We have included many of those references. Thanks.