Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2018-168-RC2, 2018
© Author(s) 2018. This work is distributed under
the Creative Commons Attribution 4.0 License.

Hydrology and Earth System Sciences
Discussions

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

MF McCabe (Referee)

matthew.mccabe@kaust.edu.sa

Received and published: 10 May 2018

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). 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.

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).

* 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.).

* 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.

* 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 hydrology-customized 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.

* 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.

* 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)

* 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?).

* 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):

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 DecisionâĂŘ-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.

* 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).

* 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.

* 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.

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.

Matthew McCabe

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