At the point of this writing, a revision has not been requested by the editor yet. This document is uploaded only to demonstrate the direction we would like to take in terms of revising this manuscript. Due to time constraint, the revision here is incomplete, immature, not-grammar-checked, and has been consented to by only a few co-authors. However, it is indeed in line with the spirit of the HESS online discussion system to give a preview of where a revision could be taken. We embrace this model and would like to communicate more than what conventional review process would have permitted.

Track change has been enabled. ***Please pay more attention to the new Sections 3 and 4****. Yellow highlight indicates author comments.*

HESS Opinions: Incubating deep-learning-powered hydrologic science advances as a community

Chaopeng Shen[1], Eric Laloy[2], Adrian Albert[3], Fi-John Chang[4], Amin Elshorbagy[5], Sangram Ganguly[6], Kuo-lin Hsu[7], Daniel Kifer[8], Zheng Fang[9], Kuai Fang[1], Dongfeng Li[9], Xiaodong Li[10], and Wen-Ping Tsai[1]

1. Civil and Environmental Engineering, Pennsylvania State University, University Park, PA 16802
2. Institute for Environment, Health and Safety, Belgian Nuclear Research Centre, Mol, Belgium
3. Energy Technologies Area, Lawrence Berkeley National Laboratory, Berkeley, CA 94720
4. Department of Bioenvironmental Systems Engineering, National Taiwan University, Taipei, 10617, Taiwan
5. Dept. of Civil, Geological, and Environmental Engineering, University of Saskatchewan, Saskatoon, Canada
6. NASA Ames Research Center/ BAER Institute, Moffett Field, CA 94035
7. Civil and Environmental Engineering, University of California, Irvine, Irvine, CA 92697
8. Computer Science and Engineering, Pennsylvania State University, University Park, PA 16802
9. Civil Engineering, University of Texas at Arlington, Arlington, TX 76013
10. State Key Laboratory of Hydraulics and Mountain River Engineering, Sichuan University, Sichuan, China

Correspondence to: Chaopeng Shen (cshen@engr.psu.edu)

Abstract. Recently, deep learning (DL) has emerged as a revolutionary and versatile tool transforming industry applications and generating new and improved capabilities for scientific discovery and model building. The adoption of DL in water science has so far been gradual, but the related fields are now ripe for breakthroughs. This paper proposes that DL-based methods can open up a viable, complementary avenue toward knowledge discovery in hydrologic sciences. In the new avenue, machine-learning algorithms present competing hypotheses that are consistent with data for scientists to further evaluate. Interrogative studies are then invoked to interpret DL models. However, hydrology presents many challenges to DL-power scientific advances, such as data limitations, model diversity and variability, and the general inexperience of the hydrologic field with

DL. The roadmap toward DL-powered scientific advances will need the coordinated effort from a large community involving scientists and citizens. Integrating process-based models with DL ones will help alleviate data limitations. The sharing of data, data pipelines, and baseline models will improve the efficiency of the community as a whole. Open competitions will greatly propel growth in hydrology and. Grass-root collaboration could overcome barriers on data science education. There are a great number of research opportunities in this new area which may stimulate advances in machine learning as well.

1. Overview

Deep learning (DL), which has gained widespread attention since 2012, is a suite of tools centering around artfully-designed large-size artificial neural networks. Compared to non-deep networks, DL is characterized by the large size to accommodate the complexities of information contained in big data, multiple levels of hidden representations, the addition of unsupervised learning units, and effective, large-scale regularization techniques. As a foundational component of modern artificial intelligence (AI), DL has made substantial strides in recent years and helped solve problems that have resisted AI for decades (LeCun et al., 2015). DL models have repeatedly been shown to outperform simpler models by large margins and generalize better to unseen instances (Schmidhuber, 2015; Shen, 2017).

Deep networks may be more robust than simpler models despite their large size, if they are regularized properly and are chosen based on validation errors in a two-stage approach (Kawaguchi et al., 2017). Effective regularization techniques include (i) early stopping: monitor the training progress on a separate validation set and stop the training once validation metrics start to deteriorate; and/or (ii) novel regularization techniques such as dropout (Srivastava et al., 2014). DL models can be easier to train than previous networks, as their architectures and new stochastic gradient techniques (Kingma and Ba, 2014) address issues like vanishing gradient (Hochreiter, 1998). Training large networks as used today was computationally implausible until scientists started to exploit the parallel processing power of graphical processing units (GPUs). Nowadays new application-specific integrated circuits have also been created to specifically tackle DL, although DL architectures are rapidly evolving.

To be expanded: more discussion of the attractive features of DL. Reduce jargons and improve readability. Citations

In contrast to many older-generation nonlinear regression and classification methods like Support Vector Machine (SVM) (Cortes and Vapnik, 1995), genetic programming (Koza, 1992), Classification and Regression Tree (CART) (Bae et al., 2010; Breiman et al., 1984) or random forest (Ho, 1995), just to name a few, deep networks are differentiable from outputs to inputs, giving them practical advantages in efficient parameter optimization via backpropagation (training). This efficiency, which is shared by some other older-generation methods like non-deep neural networks and Gaussian Processes (Snelson and Ghahramani, 2006), etc., allows DL to be used as powerful engineering and scientific design tools, whereby the often complicated effect of inputs on output variables can be estimated in a data-driven way. Moreover, the differentiable nature allows for greater success for interpolation and mild extrapolation, contributing to the strong generalization capability of DL. It has been shown that deep networks can continue to improve when the number of training instances (e.g., images) is increased

to hundreds of millions, albeit at a logarithmic rate (Sun et al., 2017). Simpler networks would have long stalled in performance prior to reaching this amount of data because they are unable to represent the complexity of the data. Lastly, like some older-generation methods, DL offers the possibility of transfer learning (Mesnil et al., 2012), where a complex deep model trained to perform a given task can be re-trained for a different but related purpose at a comparatively small computational cost. For DL, transfer learning is simple to implement: only the output layer needs to be re-trained, while the other network layers that encode a deep representation of the input data are left intact.

While DL has stimulated exciting advances in many disciplines and has become the method of choice in some areas, water sciences so far have only had a very limited set of DL applications. Despite scattered early reports of promising DL results (Fang et al., 2017; Laloy et al., 2017, 2018; Tao et al., 2016; Vandal et al., 2017; Zhang et al., 2018), water scientists seemed to have reservations about these new tools, perhaps with good reasoning. This opinion paper, endorsed by the cohort of authors, argues that there are many opportunities in water sciences where DL can help provide both stronger predictive capabilities and a complementary avenue toward scientific discovery. Readers who are less familiar with machine learning or deep learning are referred to a companion review paper (Shen, 2017) (hereafter referred to as Shen17), which provides a more comprehensive and technical background.

We first voice the opinions that elements of a complementary machine learning-based scientific discovery avenue are taking shape, and this avenue should at least be considered for problems with large data (section 2). Then, we propose several ways to accelerate this avenue (section 3). Finally, we argue that hydrology offers a unique set of challenges for DL research (section 4).

2. The emergence of a complementary avenue

We have witnessed the growth of three pillars that support a complementary research avenue utilizing deep learning: big hydrologic data, powerful machine learning algorithms, and interrogative methods to extract interpretable knowledge from the trained networks. We discuss these aspects in the following sections.

2.1. With more data, opportunities arise

The fundamental supporting factor for emerging opportunities with DL is the growth of big hydrologic data. There are ever increasing amount of hydrologic data through remote sensing (see a summary in Srinivasan, (2013)) and data compilation. Large available datasets include satellite-based data products of precipitation, surface soil moisture (Entekhabi, 2010; Jackson et al., 2016; Mecklenburg et al., 2008), vegetation states and indices, e.g., (Knyazikhin et al., 1999), and derived evapotranspiration products (Mu et al., 2011), terrestrial water storage (Wahr et al., 2006), snowcover (Hall et al., 2006), and planned mission for streamflows (Pavelsky et al., 2014), etc. On the data compilation side, there are now compilations of geologic (Gleeson et al., 2014) and soil datasets; centralized management of streamflow and groundwater data in the United

States, Europe, parts of South America and Asia, or globally for some large rivers (GRDC, 2017); water chemistry, groundwater samples and other biogeophysical datasets.

One of the 10 Big Ideas for Future Investments from U.S. National Science Foundation is "Harnessing data for 21st-century science and engineering" (NSF, 2018). With these emerging datasets, DL models can be built and trained to learn features, organizational patterns and relationships and predict outputs given new input instances. However, we are not advocating a whole transition to DL: not all problems can be suitably formulated as DL problems – they could be best tackled by specifically-designed earlier-generation methods, and for many problems, there are just not enough data to train DL-based models.

2.2. DL: A big step forward

The field of hydrology has witnessed flows and ebbs of several generations of machine learning methods in the past few decades. From regularized linear regression (Tibshirani and Tibshirani, 1994) to Support Vector Regression (Drucker et al., 1996), from genetic programming (Koza, 1992) to artificial neural networks (Chang et al., 2014; Chen et al., 2018; Hsu et al., 1995, 1997, 2002), from classification and regression tree to random forest, from Gaussian Process (Snelson and Ghahramani, 2006) to Radial Basis Function Network (Moradkhani et al., 2004), each algorithm offered useful solutions to a set of problems, but each also faces its own limitations. 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". A frequent limitation with conventional neural network study is that they are trained in a geographic region or site and typically cannot be transferred out of the training region. Large-size neural networks may be overfitted and are prohibitively expensive to train in terms of computation.

The progress brought forth by DL to the information technology industry is revolutionary (Section 4 in Shen17) and can no longer be ignored. Primary types of successful deep learning architectures include convolutional neural networks (CNN) for image recognition (Krizhevsky et al., 2012b; Ranzato et al., 2006), Long short-term memory (LSTM) (Greff et al., 2015; Hochreiter and Schmidhuber, 1997) for time series modeling, variational auto-encoders (VAE) (Kingma and Welling, 2013), and deep belief networks for pattern recognition and data (typically image but also text or sound, etc) generation (section 3.2 in Shen17). CNNs and LSTMs have earned major recognition from the industry in research spearheaded by the information technology industry. Besides these new architectures, a novel generative model concept called generative adversarial networks (GANs) has become an active area of research. The key characteristic of GANs is that they are learned by creating a competition between the actual generative model or ''generator'' and a discriminator in a zero-sum game framework (Goodfellow et al., 2014), in which these components are learned jointly. Compared to other generative models, GANs potentially offer much greater flexibility in the patterns to be generated. The power of GANs has been recognized recently in the geoscientific community, especially in machine learning research inspired by physics, where deep generative models have been used for certain complicated physical, environmental, and socio-economic systems with deep generative models (Albert et al., 2018; Laloy et al., 2018).

The evidence is mounting that when given enough data, DL can provide the unique ability to automatically extract features, sometimes better than human experts do:

- The ImageNet Challenges is an open competition to evaluate algorithms for object detection and image classification (Russakovsky et al., 2014). Topics change during each contest, and a dataset of ~14M tagged images and videos were cumulatively compiled, with convenient and uniform data access provided by the organizers. The 2010 was won by a large-scale SVM. CNNs first won this contest in 2012 (Krizhevsky et al., 2012a). Since then, and till 2017 (the last contest), the vast majority of entrants and all contest winners used CNNs, which edges out other methods by large margins (Schmidhuber, 2015).
- The IJCNN traffic sign recognition contest, which is composed of 50,000 images (48 pixels x 48 pixels), witnessed superhuman visual recognition performance from CNN-based methods (Stallkamp et al., 2011). The superhuman performance was also scored by CNNs on recognition of cancers from medical images (Yu et al., 2016).
- The TIMIT speech corpus is a dataset that holds the recordings from 630 English speakers. LSTM-based models showed a large edge over Hidden Markov Model (HMM) results (Graves et al., 2013) in recognizing the speeches. Similarly, LSTM-based methods won with large margin over all statistical approaches on keyword spotting (Indermuhle et al., 2012), optical character recognition (Breuel et al., 2013), language identification (Gonzalez-Dominguez et al., 2014), text-to-speech synthesis, social signal classification, machine translation and Chinese handwriting recognition (Schmidhuber, 2015).
- An LSTM-based speech recognition system has achieved "human parity" in conversational speech recognition on the Switchboard corpus (Xiong et al., 2016). A parallel version achieved best-known pixel-wise brain image segmentation results on the MRBrainS13 dataset (Stollenga et al., 2015). The improvement in language translation software can be witnessed by ordinary web users.
- A time series forecasting contests, Computational Intelligence in Forecasting Competition, was won by a combination of fuzzy and exponential models in 2015 when no LSTM was present, but LSTM won it in 2016 (CIF, 2016).

In sciences, DL models are quickly becoming the method of choice in analyzing data in high energy physics, chemistry, biology, astrophysics, and remote sensing (section 4.3 in Shen17), let alone medical applications such as neurosciences.

In addition to utilizing big data, DL is 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. DL was employed to achieve dynamical climate downscaling (Vandal et al., 2017), remote sensing of precipitation (Tao et al., 2017, 2018), estimate crop yield (You et al., 2017), prolong satellite-sensed soil moisture (Fang et al., 2017) and crop diseases (Pryzant et al., 2017). All these datasets are for abstract variables which can now be reliably retrieved by DL. We agree that, just like other methods, DL may eventually be replaced by newer ones, but that is not a reason to hold out on possible progress.

For revision: here, we will provide summaries of trans-disciplinary reviews (most already described in our companion review paper) for both hydrology and other disciplines. These examples, in our opinion, strongly demonstrate the advantage of DL when there are sufficient data. There aren't many big data examples in hydrology yet, but it's already showing promise. From other disciplines, the contrast is more apparent.

2.3. Network interrogative methods to enable knowledge gain from deep networks

Conventionally, neural networks were primarily used to approximate mappings between inputs and outputs. The focus was put on improving predictive accuracy. In terms of the use of neural networks in scientific research, then, there have been concerns: (1) DL and more generally machine learning (ML) are referred to as black boxes that cannot be understood by humans and thus, cannot serve to advance scientific understanding; and (2) Data-driven research lacks clearly-stated hypotheses. There has been significant pressure from inside and outside the deep learning community to make the network decisions more explainable. For example, European laws dictate that automated individual decision making which significantly influences the algorithm's users must provide a "right to explanation" where a user can ask for an explanation of an algorithmic decision (Goodman and Flaxman, 2016).

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 knowledge by deep networks from data (see literature in Section 5.2 in Shen17). Such methods include (i) attributing deep network decisions to input features or a subset of inputs. For example, for image recognition tasks, the decision of the network can be traced to some regions on the image that led to the decision (Montavon et al., 2017); (ii) transferring knowledge from deep networks to interpretable, reduced-order models. For example, a trained deep vision network can be used to train simpler models such as classification trees (Ribeiro et al., 2016); (iii) visualization of network activations, e.g., (Samek et al., 2017; Yosinski et al., 2015). For example, activations of recurrent neural networks can be visualized to show the control domain of certain cells, which explains its functioning (Karpathy et al., 2015); and (iv) problem-specific, ad-hoc analytic methods. For example, certain signals from the inputs can be added or removed to examine the impacts of such features (Alipanahi et al., 2015). Among these, (i)-(iii) are mostly developed in the computer science domain, while (iv) requires the most effort and collaboration between domain scientists and computer scientists.

Here: to give 2 concrete examples of how DL models were interpreted to help with understanding, with some overlap with Shen17 but also a new example.

2.4. The complementary research avenue

As the interrogative methods further grow, there emerges a complementary avenue toward attaining knowledge, as shown in Figure 1. The data-driven research avenue can be divided into four steps: (i) hypotheses are generated by machine learning algorithms from data; (ii) the validation step is where data withheld from training, and different from training, are employed

to evaluate the machine-learning-generated hypotheses; (iii) interpretive methods are employed to extract data-consistent and human-understandable hypotheses (described in Section 2.3); and (iv) the retained hypotheses are presented to scientists for analysis and further data collection, and the process iterates.

The classical avenue faces non-uniqueness and subjectivity. To give a concrete example, consider a classical problem of rainfall-runoff modeling. Suppose a hydrologist found that hydrologic responses in several nearby basins are different. Some basins produce flashier peaks while others have smaller peaks in summer, large seasonal fluctuation and large peak streamflows only in winter. Taking a modeling approach, the hydrologist might invoke a conceptual hydrologic model, e.g., Topmodel (Beven, 1997), however, the model results may not adequately describe the observed heterogeneity in the rainfall-runoff response. The hydrologist might hypothesize that the different behaviors are due to heterogeneity in soil texture which is not well represented in the model. The hydrologist may add in processes that represent soil spatial heterogeneity, such as modified soil pedo-transfer functions that can differentiate between the soil types in different regions. Perhaps with some parameter adjustment, this model can provide streamflow predictions that are qualitatively similar to the observations. This procedure then increases the hydrologist's confidence that the heterogeneity in soil hydraulic parameters is responsible for their different hydrologic responses. However, this improvement is not conclusive due to process equifinality: there can be alternative processes that can also result in similar outcomes, e.g., the influence of soil thickness, terrain or drainage density. The identification of potential improvement might be dependent on the hydrologist's intuition or pre-conceptions, which are nonetheless important but potentially biased. Furthermore, incorporating all the physics into the model may prove technically challenging or too time-consuming.

Compared to the classical avenue, the data-driven approach may help scientists more efficiently explore a larger set of hypotheses. Although it cannot be said that the machine learning algorithms present no human bias (because inputs are human-defined and some hyperparameters are empirically adjusted), the larger set of hypotheses presented will at least reduce that risk greatly. First, let us examine a CART-based data-driven approach. We could start with physiographic data for many basins in this region, including terrain, soil type, soil thickness, etc. We can use CART to model the process-based model's errors, which allows us to separate out the conditions under which these errors occur more frequently. We let the pattern emerge out of data without enforcing a strong human pre-conceived hypothesis. Attention must be paid to the robustness of the data mining and utilize holdout dataset or cross-validation to verify the generality of the conclusion. Data may suggest that soil thickness is the main reason for the error. Or, if data do not prefer one hypothesis over the other, then all hypotheses are equally possible and cannot be ruled out: summarized in a short phrase, "*an algorithm has no ego.*" On a practical level, this approach can more efficiently and simultaneously examine multiple competing hypotheses.

One example of such analyses was carried out in Fang and Shen, (2017) where differences in basin storage-streamflow correlations were explained by physical factors using CART, an earlier-generation data mining method (Figure 2). The data mining analysis allowed patterns to emerge, which inspired hypotheses about key factors that control the hydrologic

functioning of different systems, such as soil thickness and soil bulk density are important controls of drought recovery, while biodiversity only showed secondary importance (Schwalm et al., 2017). Scientists need to define the predictors and general model types, but they do not pose strongly constraining hypotheses about the controlling factors, and instead "let the data speak". The key to this approach is a large amount of data from which pattern emerge.

Working with DL models, we need to further resort to interrogative methods to make the results more interpretable (Figure 1 Right). For example, we can construct DL models to predict the errors of the process-based model, and then use visualization techniques to see which variable, under which condition, lead to the error. Because DL can absorb a large amount of data, it can find commonality among data as well as identify differences. Whereas CART models are limited by the amount of data and face stability problems in lower branches (data are exponentially less at lower branches), DL models may produce a more robust interpretation.

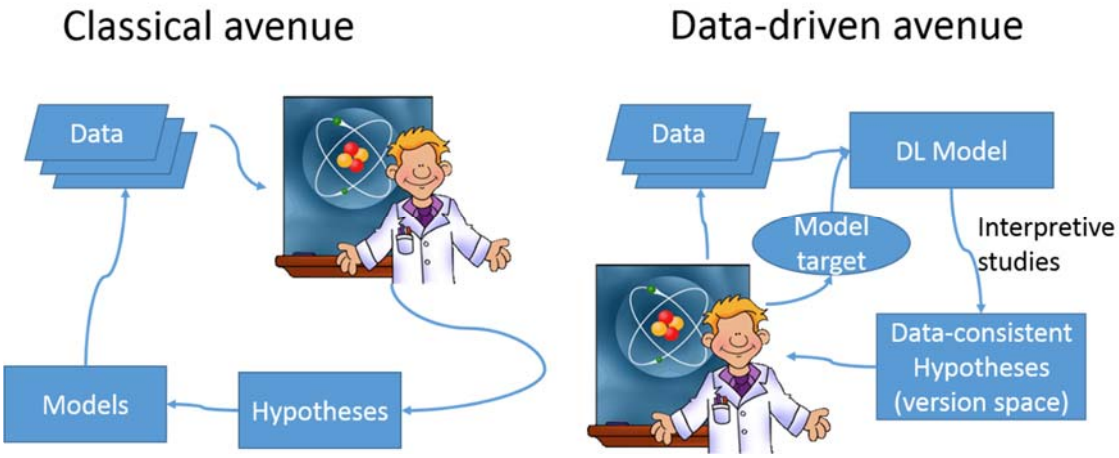

Figure 1. Comparing two alternative avenues toward gaining knowledge from data. In the classical avenue, scientists interpret data, form hypotheses, (optionally) build models to describe data and hypotheses, and then compare model results with data to affirm or reject the hypotheses. In the data-driven avenue, deep-learning models are created to learn from data to model a general, human-directed target. Then interpretive methods are employed to extract data-consistent and human-understandable hypotheses, which are presented to scientists for analysis and further data collection. There must be a hypotheses validation step where data withheld from training is used to evaluate or reject the hypotheses.

The machine learning paradigm lends us to finding "unrecognized linkages" (Wagener et al., 2010) or find complex patterns in the data that humans could not easily realize or capture. Owning to the strong capability of DL, it can better approximate the "best achievable model" (BAM) for the mapping relations between inputs and output. As such, it lends support to measuring the information content contained in the inputs about the output. Nearing et al., (2016) utilized Gaussian Process regression to approximate the BAM. DL can play similar roles and can also allow for modelling, perhaps in a more thorough way.

Outputs from the hidden layers of deep networks can now be visualized to gain insights about the transformations performed on the input data by the network (Samek et al., 2017). For image recognition tasks, one can invert the DL model to find out the parts of the inputs that led the network to make a certain decision (Mahendran and Vedaldi, 2015). There are also means to visualize outputs from recurrent networks, e.g., showing the conditions under which certain cells are activated (Karpathy et al., 2015). These visualizations can illustrate the relationships that the data-driven model has identified.

Considering the above potential benefits, the data-driven avenue should at least be considered or given an opportunity to play a role in water sciences discovery. However, this avenue may be uncomfortable to some researchers. In the classical avenue, the scientist must originate the hypotheses before constructing models; in the data-driven avenue, one needs to set up the algorithm to model a certain target. Then, the data mining/knowledge discovery process is a precursor step to the main hypotheses formation-- hypotheses cannot be generated before the data mining analysis. This feature is a natural consequence of handing part of the work to an algorithm but may cause some disarray for those who follow what has been perceived as structured scientific methods. Especially, hypotheses can no longer be unequivocally stated during the proposal stage of research.

Granted, the interrogative methods as a whole are new and time is required for them to grow. We need to note that the nascent "DL neuroscience" literature did not exist until 2015. However, if we outright reject the complementary avenue based on our habitual thinking that neural networks are black boxes, we may deny ourselves opportunities for breakthroughs.

3. Hydrology provides unique challenges and opportunities for DL

Compared to classical DL problems such as image/speech recognition that DL techniques have been applied to, hydrology has a unique set of challenges that are also research opportunities for DL. Mostly, DL research has not covered these questions extensively, but they exist across disciplines. Importantly, many of these stiff challenges cannot be sufficiently or efficiently tackled by individual research groups.~~Water scientists and computer scientists can work together to address these questions, which may lead to advances in machine learning.~~

(1) Observations in hydrology and water science, in general, are often regionally imbalanced. For example, while streamflow data are relatively dense in the United States, it is very sparse in many other parts of the world. In some parts of the world, observations have been made, but data are not made available to the public. Even for variables that can be remotely sensed, e.g., soil moisture, dense canopies often prevent uniform observations of the variable. For many hydrologic applications, there may be a dearth of observations that can be used as the supervising data. Few applications have the magnitude of data on the order of training datasets for AI tasks. A body of literature studying this problem between different geographic regions can be loosely summarized under the topic of "prediction in ungauged basins" (PUB) (Hrachowitz et al., 2013). However, PUB problems pose a significant challenge to data-driven methods~~.~~.

(2) Global change is altering the hydrologic and related cycles, and hydrologists must now make predictions in anticipation of changes, beyond previously observed ranges (Wagener et al., 2010). Especially, more frequent extremes have been observed for many parts of the world and have been projected to occur in the future. Data-driven methods often face a higher chance of failure when applied out of the range of training dataset.

(3) Hydrologic observations also tend to be incomplete in space and time, but there are multiple sources of observations focusing on different aspects of the water cycle. For example, top 5-cm surface soil moisture only reflects a very small fraction of the water cycle, but we can also observe terrestrial water storage, which is related to soil moisture. The most prevalent observation, streamflow, integrates the signal of the whole landmass. Thus, how to merge inter-related information from different sources and to improve the prediction of each other is an important question that DL have not studied extensively.

(4) Compared to standard IT applications, such as speech recognition or image recognition, water data are accompanied by a large amount of strongly heterogeneous "contextual variables" such as land use, climate, geology, and soil. Heterogeneity needs to be adequately represented without radically bloating the parameter space of the models. They covary and exert complicated controls on hydrologic responses, but we have limited knowledge of some of them, especially subsurface properties like geology. There are significant uncertainties with respect to input datasets~~,~~. In addition, these heterogeneous factors co-vary due to co-evolution (Troch et al., 2013), which makes it difficult for data-driven models to distinguish between causal and associative relationships. Especially, training with insufficient data may result in many alternative DL models that cannot be rejected.

(5) Hydrologic problems fit poorly into the template of problems that the standard network structures (Section 3.2 in Shen17) are designed for. While some direct applications such as soil moisture hindcasting (Fang et al., 2017) and precipitation retrieval from images (Tao et al., 2016) are possible, we envision many new types of problems may require customized structures. For example, catchment hydrologic problems have both spatial but static (topography and groundwater flow) and temporal (atmospheric forcing) dimensions.

(6) Because large and diverse datasets are needed, the access to datasets, their pre-processing, and appropriate formatting present practical challenges. These steps often occupy too much unnecessary time for researchers. Many of the processing tasks for images cannot be handled by a single research group. Compared to the deep learning community in AI and chemistry, etc., the machine learning in hydrology community is not sufficiently coordinated, resulting in significant waste of effort and "recreation of wheels".

(7) DL model performances can vary widely depending on model architecture, modification of network designs, training methods, use of data, data preparation, hyper-parameter setups, etc. There are a large variety of different configurations, with many options beyond what could be explored by automated algorithms. Individual research groups are often limited in only exploring part of these possibilities. Thus, it is difficult to reliably reproduce reported results and learn the advantages and

disadvantages of each model design. There are also often training "tricks" that were critical in terms of achieving the desired performance.

(8) under-coordinated community

(9) Multi-point physics challenge: needs more flexible objective function

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

Facing the above challenges, we share the vision of a community-shared roadmap toward advancing hydrologic sciences using DL. A well-coordinated community is much more efficient and powerful in resolving the abovementioned challenges. We see that several steps are crucial in this roadmap: devising ways to integrate physical knowledge, PBMs and DL, community approaches in sharing and accessing data, open and transparent model competitions, and baseline models and visualization packages. (a Figure to illustrate the roadmap)

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

To address data limitations mentioned in the last section (Points 1 through 3), we envision that an inevitable step is to more organically integrate hydrologic knowledge, process-based models, and deep learning. Process-based models, as they are derived to from underlying physics, require less data for calibration and can fill the gaps in different regions and for unobservable hydrologic processes. Given well-constructed, fundamentally-sound PBMs, they should also be able to represent the temporal changes and trends. However, because data-driven models directly target observations, they may have higher accuracy where data are available. Also, as discussed earlier, they are less prone to *a priori* model structural error. We should aim to maximally utilize the best features of each type of models.

This integration will undoubtedly be highly diversified as there can be many ways it can occur. Karpatne et al., (2017) compiled a list of approaches in the literature they collectively call "theory-guided data science" : (i) using knowledge to design data-driven model; (ii) using knowledge to initialize network states; (iii) using physical knowledge to construct priors to constrain the data-driven models; (iv) using knowledge-based constrained optimization (although this may be difficult to implement in practice); (v) using theory as regularization terms for the data-driven model, which will force the model to respect these constraints; (vi) learn hybrid models, where data-driven method is used as surrogate for certain part of the physical model. One may also impose multiple learning objectives based on the knowledge of the problem. d

This list can be further expanded to accommodate varied objectives. First, we can focus on PBM errors (difference between PBM simulation and observations). Non-deep machine learning has already shown promise in correcting PBM errors. *Abramowitz et al.* (2006) developed an ANN to predict the error in net ecosystem exchange from a land surface model, and achieved 95% reduction in annual error. More importantly, an ANN trained to correct the error at one biome completely

corrects the PBM for another, which is in a different temperature regime (Abramowitz et al., 2007). In the context of weather forecasts, machine learning methods were used to learn the patterns from past forecasting errors (Delle Monache et al., 2011, 2013). Then, through looking for similarity between the present situation and the past, error correction is advised, leading to 20% gain in performance (Junk et al., 2015). Their results suggest PBMs make structural errors that are independent of the state-variable regimes they operate in. We envision that PBMs can better resolve the impacts of regime changes, while DL can better capture state-independent error patterns and do mild state-dependent extrapolations. A co-benefits of modelling PBM error is insights about the PBM: if we are able to use interrogative methods to reverse engineer what DL has learned about these errors, it provides possible explanations to when and where our PBMs are wrong. It also provides clues as to how to fix these errors mechanistically. However, there lacks a theoretical framework for separately estimating aleatory uncertainty (resulting from data noise), and epistemic uncertainty (resulting from PBM error and training data paucity) and uncertainty due to regime-shift.-There are significant research opportunities in this regard.

Second, PBMs can provide training data for DL models, alleviating point 1 raised in Section 3. PBMs can be used to either directly create supervising data or apply perturbations to augment existing data. Furthermore, if the DL training is limited by available data, there would be many alternative DL models that could not be rejected (point 4 in the last section). Some of these alternative DL models generate unphysical outputs. Providing PBM simulations as either training data or regularization terms help to nudge DL models to generate physically meaningful outputs.

Here, two additional ways of PBM-DL integration will be proposed.

In summary, there is substantial potential in combining the benefits of DL and PBMs. There are myriad possible approaches, yet guiding theories are lacking. On top of these alternatives, different hydrologic problems, e.g., soil moisture or streamflow, have different system properties. The advantages and disadvantages of these approaches could be systematically and efficiently evaluated in community-coordinated fashion.

4.2. Community-coordinated hydrologic modeling competitions to pursue both performance and explainability

As mentioned in Section 4.1 and points 6 and 7 in Section 3, there are many possible approaches and many alternative model structures. In the light of these challenges, ~~W~~we argue that open, fast and standardized competitions are a very effective way of accelerating the progress in an area. In addition to commonly employed metrics, we can formulate competitions that are evaluated based on the attainment of understanding.

As discussed earlier, the effectiveness of competitions is best demonstrated in the community-coordinated challenges in computer science. These competitions have strongly propelled advances in artificial intelligence. New methods can be evaluated objectively and disseminated rapidly~~, with reduced subjectivity in the evaluation~~. Because the problems are standardized, they remove significant variability in terms of data sources and pre-processing. In the case of deep learning, DL models have emerged as a dominant force in almost every contest where it was applicable since 2012. Despite substantial

manual efforts spent on earlier methods such as SVM and Hidden Markov Model (HMM), deep neural networks repeatedly show advantages. Via these competitions, the community can quickly learn advantages and disadvantages of alternative ~~some~~ network designs. They encourage results to be reproducible and comparable. The advantages and disadvantages of different methods can be thoroughly explored and laterally compared. ~~However, these results do not suggest other statistical methods have no value. Rather, data limitations and various constraints often make simpler methods valuable. However, an undeniable and increasingly apparent trend is that deep learning shows unrivaled predictive performance~~.

We envision multi-faceted hydrologic modeling competitions where various models ranging from process-based ones to deep learning ones are evaluated and compared. The coordinators will provide a set of standard atmospheric forcings, landscape characteristics, and observed variables. The participants should submit results driven by the standard inputs, but they may optionally also use their own inputs. Target observations may be soil moisture, streamflow records and/or groundwater levels. Importantly, the evaluation criteria include not only performance-type criteria such as model efficiency coefficients and bias but also **qualitative/explanatory** ones such as explanations for control variables and model errors. Over-simplified or poorly-constructed models may provide more accessible explanations, but they might be misleading because the models may be overfitted to a given situation. Their simplicity may also constrain their ability to digest large datasets as a way of reducing uncertainty. Multi-faceted competitions allow us to also identify a "Pareto front" of explainability and performance and help rule out "false explanations". The objective of the competition is not only to seek the best simulation performance, but also those methods that offer deeper insight into hydrologic dynamics.

Another important value of competitions is that organizers will provide a standard input dataset and well-defined tasks, which greatly save the community resources and effort, so that participants can focus on the modeling aspects. A substantial amount of effort is required to establish such a dataset, which may only be possible under a specifically designed project. However, any effort in creating the dataset will return great value to the community.

4.3. Community-shared data sources and processing pipelines

The main approach to address the major obstacle of data limitation is to increase our data repositories. In addition, facilitating access to datasets is also an important aspect of increasing our data availability. Data collection can be greatly enhanced by centralized data compilation, a task many institutions are already undertaking. For example, the Consortium of Universities for the Advancement of Hydrologic Science, Inc. (CUAHSI) hosts large amounts of hydrologic data. For another example, in 2015, a project called Collaborative Research Actions (Endo et al., 2015) was proposed in Belmont Forum, which is a group of the world's major and emerging funders of global environmental change research. Many scientists from different countries join the project and focus on the same issue, Food-Energy-Water Nexus. They shared their data (heterogeneous data) and research results from different regions.

The database organizers could help format data in a way that facilitates data mining and deep learning. However, it is unlikely that all data can be stored in one location, considering the volume of high-resolution remote sensing data. This coordination would require consulting with data scientists when designing the infrastructure. In addition, besides providing data, a concurrent role that databases can take is also to provide more channels to share experiences, scholarly discussions, and debates along with the generation of data.

Another important area where deep learning is expected to deliver significant value is the analysis of big and sub-research-quality data such as those collected by citizen scientists. A valuable feature of water sciences is that they are accessible to ordinary people. Citizen scientists could help gather data about precipitation, temperature, humidity, soil moisture, river stage, and potentially groundwater levels. These quantities can be measured by inexpensive instruments like pressure gauges and moisture sensors. To add: new remote sensing methods like CubeSat and drones. To cite McCabe et al. HESS 2017, etc.. Volunteer scientists can also be requested for results in an active learning framework (Settles, 2012), i.e., they can be queried for more data for instances that can best reduce the uncertainty of the predictions. Crowd-sourced data have played roles in deep learning research (Huang et al., 2016; Izadinia et al., 2015), even though there are problems related to data quality. An important co-benefits of involving citizen scientists is the education and outreach to the public. The active engagement is much more effective when the public has a stake in the research outcomes.

4.4. Develop a base suite of shared models, interpretation and visualization software

To be expanded: these shared models, analogical to GoogleNet, etc., could greatly facilitate newcomers in getting started. In addition, they improve reproducibility and the effectiveness of comparisons. The interpretation and visualization effort seem fragmented and adhoc. Compiling and collecting them into community-shared resources will greatly improve our growth.

4.5. Education

To be expanded: a huge barrier is our educational background. There is very little preparation for big data. Grass-root collaboration could overcome barriers e data science education.

5. Concluding remarks

In this opinion paper, we argue that scientists ought to give thoughts to a complementary research avenue, where DL-power data mining is used to generate hypotheses, which ~~are subsequently tested~~c. In the past there may have been strong reservations toward black-box machine learning algorithms. Significant efforts have been put in the interpretation and understanding of deep learning networks, and hydrologists have the opportunity to push research forward in this regard. Progress in hydrology

and other disciplines show that there is substantial promise in incorporating DL into hydrologists war chest. However, challenges such as data limitation and model variability demand a community-coordinated approach.

We have also argued for open hydrologic competitions that emphasize both performance and explainability. These competitions, along with shared data, DL models and data pipelines, will greatly improve the growth of the field as a whole. ~~DL has powered breakthroughs in other disciplines. We argue water sciences~~Hydrologists should make use of the ~~big data and~~potential of citizen science ~~potential~~, and exploit DL as a valuable tool toward scientific discovery ~~in water related fields~~.

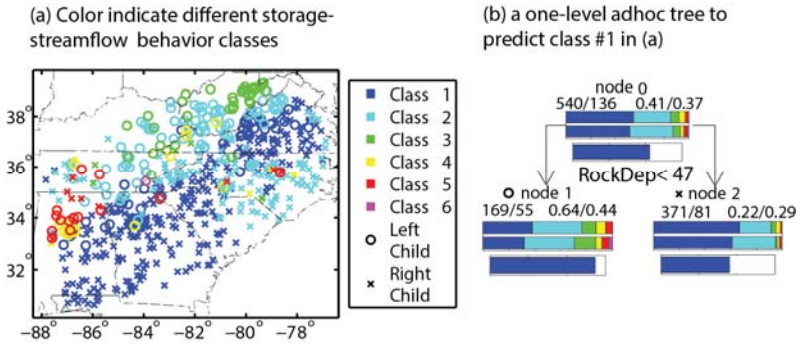

Figure 2. (adapted from Fang and Shen 2017. Reprint permission obtained). We calculated storage-streamflow correlation patterns over continental United States (CONUS) and divided small or mesoscale basins into multiple classes. We studied what physical factors most cleanly separate different correlation patterns. In this case, what separates the blue class (storage and streamflow are highly correlated across all flow regimes) and the green class turned out to be soil thickness. It suggests the blue basins in the south has good correlation because they have thick soils, which facilitates infiltration, water storage, and groundwater-dominated streamflow.

References

Albert, A., Strano, E., Kaur, J. and Gonzalez, M.: Modeling urbanization patterns with generative adversarial networks, arXiv:1801.02710 [online] Available from: http://arxiv.org/abs/1801.02710 (Accessed 24 March 2018), 2018.

Alipanahi, B., Delong, A., Weirauch, M. T. and Frey, B. J.: Predicting the sequence specificities of DNA- and RNA-binding proteins by deep learning, Nat. Biotechnol., 33(8), 831–838, doi:10.1038/nbt.3300, 2015.

Bae, H.-K., Olson, B. H., Hsu, K.-L. and Sorooshian, S.: Classification and regression tree (CART) analysis for indicator bacterial concentration prediction for a Californian coastal area, Water Sci. Technol., 61(2), 545, doi:10.2166/wst.2010.842, 2010.

Beven, K.: Topmodel : A Critique, Hydrol. Process., 11(December 1996), 1069–1085, 1997.

Breiman, L., Friedman, J., Olshen, R. and Stone, C.: Classification and Regression Trees, CRC Press., 1984.

Breuel, T. M., Ul-Hasan, A., Al-Azawi, M. A. and Shafait, F.: High-Performance OCR for Printed English and Fraktur Using LSTM Networks, in 2013 12th International Conference on Document Analysis and Recognition, pp. 683–687, IEEE., 2013.

Chang, L.-C., Shen, H.-Y. and Chang, F.-J.: Regional flood inundation nowcast using hybrid SOM and dynamic neural networks, J. Hydrol., 519, 476–489, doi:10.1016/J.JHYDROL.2014.07.036, 2014.

Chen, I.-T., Chang, L.-C. and Chang, F.-J.: Exploring the spatio-temporal interrelation between groundwater and surface water by using the self-organizing maps, J. Hydrol., 556, 131–142, doi:10.1016/J.JHYDROL.2017.10.015, 2018.

CIF: Results, Int. Time Ser. Forecast. Compet. - Comput. Intell. Forecast. [online] Available from: http://irafm.osu.cz/cif/main.php?c=Static&page=results (Accessed 24 March 2018), 2016.

Cortes, C. and Vapnik, V.: Support-vector networks, Mach. Learn., 20(3), 273–297, doi:10.1007/BF00994018, 1995.

Drucker, H., Burges, C. J. C., Kaufman, L., Smola, A. and Vapnik, V.: Support vector regression machines, Proc. 9th Int. Conf. Neural Inf. Process. Syst., 155–161 [online] Available from: https://dl.acm.org/citation.cfm?id=2999003 (Accessed 5 January 2018), 1996.

Endo, A., Burnett, K., Orencio, P., Kumazawa, T., Wada, C., Ishii, A., Tsurita, I. and Taniguchi, M.: Methods of the Water-Energy-Food Nexus, Water, 7(10), 5806–5830, doi:10.3390/w7105806, 2015.

Entekhabi, D.: The Soil Moisture Active Passive (SMAP) mission, Proc. IEEE, 98(5), 704–716, doi:10.1109/JPROC.2010.2043918, 2010.

Fang, K. and Shen, C.: Full-flow-regime storage-streamflow correlation patterns provide insights into hydrologic functioning over the continental US, Water Resour. Res., doi:10.1002/2016WR020283, 2017.

Fang, K., Shen, C., Kifer, D. and Yang, X.: Prolongation of SMAP to Spatio-temporally Seamless Coverage of Continental US Using a Deep Learning Neural Network, Geophys. Res. Lett., doi:10.1002/2017GL075619, 2017.

Gleeson, T., Moosdorf, N., Hartmann, J. and van Beek, L. P. H.: A glimpse beneath earth's surface: GLobal HYdrogeology MaPS (GLHYMPS) of permeability and porosity, Geophys. Res. Lett., 41(11), 3891–3898, doi:10.1002/2014GL059856, 2014.

Gonzalez-Dominguez, J., Lopez-Moreno, I., Sak, H., Gonzalez-Rodriguez, J. and Moreno, P. J.: Automatic Language Identification using Long Short-Term Memory Recurrent Neural Networks, in Interspeech 2014., 2014.

Goodfellow, I. J., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A. and Bengio, Y.: Generative Adversarial Networks, in Proceedings of the 27th International Conference on Neural Information Processing Systems (NIPS'14). [online] Available from: http://arxiv.org/abs/1406.2661 (Accessed 25 February 2017), 2014.

Goodman, B. and Flaxman, S.: European Union regulations on algorithmic decision-making and a "right to explanation", arXiv:1606.08813 [online] Available from: http://arxiv.org/abs/1606.08813 (Accessed 7 February 2018), 2016.

Graves, A., Mohamed, A. and Hinton, G.: Speech Recognition with Deep Recurrent Neural Networks, in ICASSP 2013., 2013.

GRDC: River Discharge Data, Glob. Runoff Data Cent. [online] Available from: http://www.bafg.de/GRDC/EN/02_srvcs/21_tmsrs/riverdischarge_node.html (Accessed 28 July 2017), 2017.

Greff, K., Srivastava, R. K., Koutník, J., Steunebrink, B. R. and Schmidhuber, J.: LSTM: A Search Space Odyssey, http://arxiv.org/abs/1503.04069 [online] Available from: http://arxiv.org/abs/1503.04069 (Accessed 18 July 2016), 2015.

Hall, D. K., Riggs, G. A. and Salomonson., V. V.: MODIS/Terra Snow Cover 5-Min L2 Swath 500m. Version 5., Boulder, Colorado USA., 2006.

Ho, T. K.: Random decision forests, in Proceeding ICDAR '95 Proceedings of the Third International Conference on Document Analysis and Recognition., 1995.

Hochreiter, S.: The Vanishing Gradient Problem During Learning Recurrent Neural Nets and Problem Solutions, Int. J. Uncertainty, Fuzziness Knowledge-Based Syst., 6(2), 107–116, doi:10.1142/S0218488598000094, 1998.

Hochreiter, S. and Schmidhuber, J.: Long Short-Term Memory, Neural Comput., 9(8), 1735–1780, doi:10.1162/neco.1997.9.8.1735, 1997.

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., Arheimer, B., Blume, T., Clark, M. P., Ehret, U., Fenicia, F., Freer, J. E., Gelfan, A., Gupta, H. V., Hughes, D. A., Hut, R. W., Montanari, A., Pande, S., Tetzlaff, D., Troch, P. A., Uhlenbrook, S., Wagener, T., Winsemius, H. C., Woods, R. A., Zehe, E. and Cudennec, C.: A decade of Predictions in Ungauged Basins (PUB)—a review, Hydrol. Sci. J., 58(6), 1198–1255, doi:10.1080/02626667.2013.803183, 2013.

Hsu, K., Gao, X., Sorooshian, S., Gupta, H. V., Hsu, K., Gao, X., Sorooshian, S. and Gupta, H. V.: Precipitation Estimation from Remotely Sensed Information Using Artificial Neural Networks, J. Appl. Meteorol., 36(9), 1176–1190, doi:10.1175/1520-0450(1997)036<1176:PEFRSI>2.0.CO;2, 1997.

Hsu, K., Gupta, H. V., Gao, X., Sorooshian, S. and Imam, B.: Self-organizing linear output map (SOLO): An artificial neural network suitable for hydrologic modeling and analysis, Water Resour. Res., 38(12), 38-1-38–17, doi:10.1029/2001WR000795, 2002.

Hsu, K., Gupta, H. V. and Sorooshian, S.: Artificial Neural Network Modeling of the Rainfall-Runoff Process, Water Resour. Res., 31(10), 2517–2530, doi:10.1029/95WR01955, 1995.

Huang, W., He, D., Yang, X., Zhou, Z., Kifer, D. and Giles, C. L.: Detecting Arbitrary Oriented Text in the Wild with a Visual Attention Model, in Proceedings of the 2016 ACM on Multimedia Conference - MM '16, pp. 551–555, ACM Press, New York, New York, USA., 2016.

Indermuhle, E., Frinken, V. and Bunke, H.: Mode Detection in Online Handwritten Documents Using BLSTM Neural Networks, in 2012 International Conference on Frontiers in Handwriting Recognition, pp. 302–307, IEEE., 2012.

Izadinia, H., Russell, B. C., Farhadi, A., Hoffman, M. D. and Hertzmann, A.: Deep Classifiers from Image Tags in the Wild, in Proceedings of the 2015 Workshop on Community-Organized Multimodal Mining: Opportunities for Novel Solutions, pp. 13–18, ACM., 2015.

Jackson, T., O'Neill, P., Njoku, E., Chan, S., Bindlish, R., Colliander, A., Chen, F., Burgin, M., Dunbar, S., Piepmeier, J., Cosh, M., Caldwell, T., Walker, J., Wu, X., Berg, A., Rowlandson, T., Pacheco, A., McNairn, H., Thibeault, M., Martínez-Fernández9, J., González-Zamora, Á., Seyfried10, M., Bosch, D.,

Starks, P., Goodrich, D., Prueger, J., Su, Z., van der Velde, R., Asanuma, J., Palecki, M., Small, E., Zreda, M., Calvet, J., Crow, W., Kerr, Y., Yueh, S. and Entekhabi, D.: Soil Moisture Active Passive (SMAP) Project Calibration and Validation for the L2/3_SM_P Version 3 Data Products, SMAP Proj. JPL D-93720 [online] Available from: http://nsidc.org/data/docs/daac/smap/sp_l2_smp/pdfs/L2SMP_validated_assess_rpt_rel2_v10a_final.pdf (Accessed 27 July 2017), 2016.

Karpathy, A., Johnson, J. and Fei-Fei, L.: Visualizing and Understanding Recurrent Networks, in ICLR 2016 Workshop. [online] Available from: http://arxiv.org/abs/1506.02078 (Accessed 7 November 2016), 2015.

Kawaguchi, K., Kaelbling, L. P. and Bengio, Y.: Generalization in Deep Learning, arXiv:1710.05468 [online] Available from: http://arxiv.org/abs/1710.05468 (Accessed 12 March 2018), 2017.

Kingma, D. P. and Ba, J.: Adam: A Method for Stochastic Optimization, in 3rd International Conference for Learning Representations, San Diego, CA. [online] Available from: http://arxiv.org/abs/1412.6980 (Accessed 30 March 2018), 2014.

Kingma, D. P. and Welling, M.: Auto-Encoding Variational Bayes, in Proceedings of the 2014 International Conference on Learning Representations (ICLR). [online] Available from: http://arxiv.org/abs/1312.6114 (Accessed 24 March 2018), 2013.

Knyazikhin, Y., Glassy, J., Privette, J. L., Tian, Y., Lotsch, A., Zhang, Y., Wang, Y., Morisette, J. T., P.Votava, Myneni, R. B., Nemani, R. R. and Running, S. W.: MODIS Leaf Area Index (LAI) and Fraction of Photosynthetically Active Radiation Absorbed by Vegetation (FPAR) Product (MOD15) Algorithm Theoretical Basis Document, http://eospso.gsfc.nasa.gov/atbd/modistables.html, 1999., 1999.

Koza, J. R.: Genetic Programming: on the Programming of Computers by Means of Natural Selection, MIT Press., 1992.

Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet classification with deep convolutional neural networks, Proc. 25th Int. Conf. Neural Inf. Process. Syst. - Vol. 1, 1097–1105 [online] Available from: https://dl.acm.org/citation.cfm?id=2999257 (Accessed 10 March 2018a), 2012.

Krizhevsky, A., Sutskever, I. and Hinton, G. E.: ImageNet Classification with Deep Convolutional Neural Networks, in Advances in Neural Information Processing Systems 25, pp. 1097–1105. [online] Available from: https://papers.nips.cc/paper/4824-imagenet-classification-with-deep-convolutional-neural-networks (Accessed 30 March 2018b), 2012.

Laloy, E., Hérault, R., Jacques, D. and Linde, N.: Training-Image Based Geostatistical Inversion Using a Spatial Generative Adversarial Neural Network, Water Resour. Res., 54(1), 381–406, doi:10.1002/2017WR022148, 2018.

Laloy, E., Hérault, R., Lee, J., Jacques, D. and Linde, N.: Inversion using a new low-dimensional representation of complex binary geological media based on a deep neural network, Adv. Water Resour., 110, 387–405, doi:10.1016/J.ADVWATRES.2017.09.029, 2017.

LeCun, Y., Bengio, Y. and Hinton, G.: Deep learning, Nature, 521(7553), 436–444, doi:10.1038/nature14539, 2015.

Liu, Y., Racah, E., Prabhat, Correa, J., Khosrowshahi, A., Lavers, D., Kunkel, K., Wehner, M. and Collins, W.: Application of Deep Convolutional Neural Networks for Detecting Extreme Weather in Climate Datasets, in ACM SIGKDD 2016 Conference on Knowledge Discovery & Data Mining. [online] Available from: http://arxiv.org/abs/1605.01156 (Accessed 21 October 2016), 2016.

Mahendran, A. and Vedaldi, A.: Understanding deep image representations by inverting them, in 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 5188–5196, IEEE., 2015.

Matsuoka, D., Nakano, M., Daisuke Sugiyama and Uchida, S.: Detecting Precursors of Tropical Cyclone using Deep Neural Networks, in The 7th International Workshop on Climate Informatics: CI 2017., 2017.

Mecklenburg, S., Kerr, Y., Font, J. and Hahne, A.: The Soil Moisture and Ocean Salinity Mission - An Overview, in IGARSS 2008 - 2008 IEEE International Geoscience and Remote Sensing Symposium, p. IV-938-IV-941, IEEE., 2008.

Mesnil, G., Dauphin, Y., Glorot, X., Rifai, S., Bengio, Y., Goodfellow, I., Lavoie, E., Muller, X., Desjardins, G., Warde-Farley, D., Vincent, P., Courville, A. and Bergstra, J.: Unsupervised and Transfer Learning Challenge: a Deep Learning Approach, in Proceedings of ICML Workshop on Unsupervised and Transfer Learning, vol. 27, edited by I. Guyon, G. Dror, V. Lemaire, G. Taylor, and D. Silver, pp. 97–110, PMLR, Bellevue, Washington, USA. [online] Available from: http://proceedings.mlr.press/v27/mesnil12a.html, 2012.

Montavon, G., Samek, W. and Müller, K.-R.: Methods for Interpreting and Understanding Deep Neural Networks, Digit. Signal Process., doi:10.1016/J.DSP.2017.10.011, 2017.

Moradkhani, H., Hsu, K., Gupta, H. V. and Sorooshian, S.: Improved streamflow forecasting using self-organizing radial basis function artificial neural networks, J. Hydrol., 295(1–4), 246–262, doi:10.1016/J.JHYDROL.2004.03.027, 2004.

Mosser, L., Dubrule, O. and Blunt, M. J.: Reconstruction of three-dimensional porous media using generative adversarial neural networks, Phys. Rev. E, 96(4), 43309, doi:10.1103/PhysRevE.96.043309, 2017.

Mu, Q., Zhao, M. and Running, S. W.: Improvements to a MODIS global terrestrial evapotranspiration algorithm, Remote Sens. Environ., 115(8), 1781–1800, doi:10.1016/j.rse.2011.02.019, 2011.

Nearing, G. S., Mocko, D. M., Peters-Lidard, C. D., Kumar, S. V., Xia, Y., Nearing, G. S., Mocko, D. M., Peters-Lidard, C. D., Kumar, S. V. and Xia, Y.: Benchmarking NLDAS-2 Soil Moisture and Evapotranspiration to Separate Uncertainty Contributions, J. Hydrometeorol., 17(3), 745–759, doi:10.1175/JHM-D-15-0063.1, 2016.

NSF: NSF's 10 Big Ideas, Natl. Sci. Found. Spec. Rep. [online] Available from: https://www.nsf.gov/news/special_reports/big_ideas/ (Accessed 25 February 2018), 2018.

Pavelsky, T. M., Durand, M. T., Andreadis, K. M., Beighley, R. E., Paiva, R. C. D., Allen, G. H. and Miller, Z. F.: Assessing the potential global extent of SWOT river discharge observations, J. Hydrol., 519, 1516–1525, doi:10.1016/j.jhydrol.2014.08.044, 2014.

Pryzant, R., Ermon, S. and Lobell, D.: Monitoring Ethiopian Wheat Fungus with Satellite Imagery and Deep Feature Learning, in 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), pp. 1524–1532, IEEE., 2017.

Ranzato, M., Poultney, C., Chopra, S. and LeCun, Y.: Efficient learning of sparse representations with an energy-based model, Proc. 19th Int. Conf. Neural Inf. Process. Syst., 1137–1144 [online] Available from: https://dl.acm.org/citation.cfm?id=2976599 (Accessed 30 March 2018), 2006.

Ribeiro, M. T., Singh, S. and Guestrin, C.: "Why Should I Trust You?": Explaining the Predictions of Any Classifier, in Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16, pp. 1135–1144, ACM Press, New York, New York, USA., 2016.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., Berg, A. C. and Fei-Fei, L.: ImageNet Large Scale Visual Recognition Challenge, arXiv:1409.0575 [online] Available from: http://arxiv.org/abs/1409.0575 (Accessed 24 March 2018), 2014.

Samek, W., Binder, A., Montavon, G., Lapuschkin, S. and Muller, K.-R.: Evaluating the Visualization of What a Deep Neural Network Has Learned, IEEE Trans. Neural Networks Learn. Syst., 28(11), 2660–2673, doi:10.1109/TNNLS.2016.2599820, 2017.

Schmidhuber, J.: Deep learning in neural networks: An overview, Neural Networks, 61, 85–117, doi:10.1016/j.neunet.2014.09.003, 2015.

Schwalm, C. R., Anderegg, W. R. L., Michalak, A. M., Fisher, J. B., Biondi, F., Koch, G., Litvak, M., Ogle, K., Shaw, J. D., Wolf, A., Huntzinger, D. N., Schaefer, K., Cook, R., Wei, Y., Fang, Y., Hayes, D., Huang, M., Jain, A. and Tian, H.: Global patterns of drought recovery, Nature, 548(7666), 202–205, doi:10.1038/nature23021, 2017.

Settles, B.: Active Learning, in Synthesis lectures on artificial intelligence and machine learning, edited by R. J. Brachman, W. W. Cohen, and T. G. Dietterich, Norgan & Claypool., 2012.

Shen, C.: A trans-disciplinary review of deep learning research for water resources scientists, arXiv:1712.02162 [online] Available from: http://arxiv.org/abs/1712.02162 (Accessed 3 January 2018), 2017.

Snelson, E. and Ghahramani, Z.: Sparse Gaussian Processes using Pseudo-inputs, Adv. Neural Inf. Process. Syst., 18, 1257--1264, 2006.

Srinivasan, M.: Hydrology from space: NASA's satellites supporing water resources applications, Water Forum III Droughts Other Extrem. Weather Events [online] Available from: http://www.jsg.utexas.edu/ciess/files/Srinivasanetal_TWF_Oct14_Final.pdf (Accessed 12 July 2016), 2013.

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I. and Salakhutdinov, R.: Dropout: A Simple Way to Prevent Neural Networks from Overfitting, J. Mach. Learn. Res., 15, 1929–1958 [online] Available from: http://jmlr.org/papers/v15/srivastava14a.html (Accessed 28 November 2015), 2014.

Stallkamp, J., Schlipsing, M., Salmen, J. and Igel, C.: The German Traffic Sign Recognition Benchmark: A multi-class classification competition, in The 2011 International Joint Conference on Neural Networks, pp. 1453–1460, IEEE., 2011.

Stollenga, M. F., Byeon, W., Liwicki, M. and Schmidhuber, J.: Parallel multi-dimensional LSTM, with application to fast biomedical volumetric image segmentation, Proc. 28th Int. Conf. Neural Inf. Process. Syst. - Vol. 2, 2998–3006 [online] Available from: https://dl.acm.org/citation.cfm?id=2969574 (Accessed 16 October 2017), 2015.

Sun, C., Shrivastava, A., Singh, S. and Gupta, A.: Revisiting Unreasonable Effectiveness of Data in Deep Learning Era, in ICCV 2017. [online] Available from: http://arxiv.org/abs/1707.02968 (Accessed 1 December 2017), 2017.

Tao, Y., Gao, X., Hsu, K., Sorooshian, S. and Ihler, A.: A Deep Neural Network Modeling Framework to Reduce Bias in Satellite Precipitation Products, J. Hydrometeorol., doi:JHM-D-15-0075.1, 2016.

Tao, Y., Gao, X., Ihler, A., Sorooshian, S., Hsu, K., Tao, Y., Gao, X., Ihler, A., Sorooshian, S. and Hsu, K.: Precipitation Identification with Bispectral Satellite Information Using Deep Learning Approaches, J. Hydrometeorol., 18(5), 1271–1283, doi:10.1175/JHM-D-16-0176.1, 2017.

Tao, Y., Hsu, K., Ihler, A., Gao, X., Sorooshian, S., Tao, Y., Hsu, K., Ihler, A., Gao, X. and Sorooshian, S.: A Two-Stage Deep Neural Network Framework for Precipitation Estimation from Bispectral Satellite Information, J. Hydrometeorol., 19(2), 393–408, doi:10.1175/JHM-D-17-0077.1, 2018.

Tibshirani, R. and Tibshirani, R.: Regression Shrinkage and Selection Via the Lasso, J. R. Stat. Soc. Ser. B, 58, 267--288 [online] Available from: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.35.7574 (Accessed 4 August 2017), 1994.

Troch, P. A., Carrillo, G., Sivapalan, M., Wagener, T. and Sawicz, K.: Climate-vegetation-soil interactions and long-term hydrologic partitioning: signatures of catchment co-evolution, Hydrol. Earth Syst. Sci., 17(6), 2209–2217, doi:10.5194/hess-17-2209-2013, 2013.

Vandal, T., Kodra, E., Ganguly, S., Michaelis, A., Nemani, R. and Ganguly, A. R.: DeepSD: Generating High Resolution Climate Change Projections through Single Image Super-Resolution, in 23rd ACM SIGKDD Conference on Knowledge Discovery and Data Mining. [online] Available from: http://arxiv.org/abs/1703.03126 (Accessed 2 December 2017), 2017.

Wagener, T., Sivapalan, M., Troch, P. a., McGlynn, B. L., Harman, C. J., Gupta, H. V., Kumar, P., Rao, P. S. C., Basu, N. B. and Wilson, J. S.: The future of hydrology: An evolving science for a changing world, Water Resour. Res., 46(5), 1–10, doi:10.1029/2009WR008906, 2010.

Wahr, J., Swenson, S. and Velicogna, I.: Accuracy of GRACE mass estimates, Geophys. Res. Lett., 33(6), L06401, doi:10.1029/2005GL025305, 2006.

Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., Yu, D. and Zweig, G.: Achieving Human Parity in Conversational Speech Recognition, Microsoft Res. Tech. Rep. MSR-TR-2016-71. arXiv1610.05256 [online] Available from: http://arxiv.org/abs/1610.05256 (Accessed 24 March 2018), 2016.

Yosinski, J., Clune, J., Nguyen, A., Fuchs, T. and Lipson, H.: Understanding Neural Networks Through Deep Visualization, in Deep Learning Workshop, 31 st International Conference on Machine Learning, Lille, France. [online] Available from: http://arxiv.org/abs/1506.06579 (Accessed 19 November 2017), 2015.

You, J., Li, X., Low, M., Lobell, D. and Ermon, S.: Deep Gaussian Process for Crop Yield Prediction Based on Remote Sensing Data, in Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence (AAAI-17)., 2017.

Yu, K.-H., Zhang, C., Berry, G. J., Altman, R. B., Ré, C., Rubin, D. L. and Snyder, M.: Predicting non-small cell lung cancer prognosis by fully automated microscopic pathology image features, Nat. Commun., 7, 12474, doi:10.1038/ncomms12474, 2016.

Zhang, D., Lindholm, G. and Ratnaweera, H.: Use long short-term memory to enhance Internet of Things for combined sewer overflow monitoring, J. Hydrol., 556, 409–418, doi:10.1016/J.JHYDROL.2017.11.018, 2018.