Geophysical Research Letters

RESEARCH LETTER
10.1029/2020GL088229

Improving AI System Awareness of Geoscience Knowledge: Symbiotic Integration of Physical Approaches and Deep Learning

Shijie Jiang[1,2], **Yi Zheng**[1,3], **and Dimitri Solomatine**[4,5,6]

[1]School of Environmental Science and Engineering, Southern University of Science and Technology, Shenzhen, China, [2]Department of Civil and Environmental Engineering, National University of Singapore, Singapore, [3]Shenzhen Municipal Engineering Lab of Environmental IoT Technologies, Southern University of Science and Technology, Shenzhen, China, [4]Department of Hydroinformatics and Socio-Technical Innovation, IHE Delft Institute for Water Education, Delft, Netherlands, [5]Department of Water Management, Delft University of Technology, Delft, Netherlands, [6]Water Problems Institute, Russian Academy of Sciences, Moscow, Russia

Key Points:
- A novel approach to representing geosystem dynamics via a recurrent neural network within deep learning architectures is proposed
- The physics-aware AI system exhibits robust transferability and good intelligence for inferring unobserved processes in runoff modeling
- The hydrology-aware DL model can be an intelligent parameterization module for the encoded hydrologic model in cross-region applications

Supporting Information:
- Supporting Information S1

Correspondence to:
Y. Zheng,
zhengy@sustech.edu.cn

Citation:
Jiang, S., Zheng, Y., & Solomatine, D. (2020). Improving AI system awareness of geoscience knowledge: Symbiotic integration of physical approaches and deep learning. *Geophysical Research Letters*, *46*, e2020GL088229. https://doi.org/10.1029/2020GL088229

Received 1 APR 2020
Accepted 3 JUN 2020
Accepted article online 9 JUN 2020

Abstract Modeling dynamic geophysical phenomena is at the core of Earth and environmental studies. The geoscientific community relying mainly on physical representations may want to consider much deeper adoption of artificial intelligence (AI) instruments in the context of AI's global success and emergence of big Earth data. A new perspective of using hybrid physics-AI approaches is a grand vision, but actualizing such approaches remains an open question in geoscience. This study develops a general approach to improving AI geoscientific awareness, wherein physical approaches such as temporal dynamic geoscientific models are included as special recurrent neural layers in a deep learning architecture. The illustrative case of runoff modeling across the conterminous United States demonstrates that the physics-aware DL model has enhanced prediction accuracy, robust transferability, and good intelligence for inferring unobserved processes. This study represents a firm step toward realizing the vision of tackling Earth system challenges by physics-AI integration.

Plain Language Summary Artificial intelligence (AI) learns and makes inferences from experience and resembles the way untaught humans learn. If scientists can manage to teach AI the physical rules of the world, the "educated" AI may be more intelligent in deductions. However, how to implant elements of physical representations into an AI system effectively and directly remains an open question. This study proposes a general framework and applicable solutions to this challenge in the context of Earth science. The novel framework has a specially structured design for an AI system to "memorize" physical rules behind system dynamics (i.e., how a geosystem evolves with time). Following this framework, we developed a hydrology-aware deep learning model to simulate/predict runoff in 569 catchments across the conterminous United States. The results show that after "learning" a hydrologic model, the AI system has enhanced prediction accuracy and good intelligence to deal with unfamiliar regions and infer unobserved processes. The potential of AI for in-depth information mining, in return, fills the knowledge gap existing in physical approaches. The symbiotic integration of physical approaches and deep learning represents a promising solution to improve AI system awareness of geoscience knowledge.

1. Introduction

As a cornerstone of geosciences, physical approaches have achieved notable success in explaining and predicting the state changes in a geosystem (Bauer et al., 2015; Eyring et al., 2019), whereas nowadays they have to confront the development of artificial intelligence (AI), especially the recent development of deep learning (DL). In the context of the emergence of big Earth data (Yang et al., 2016), DL evolved into a budding tool for making scientific predictions independent of physical principles and its application is receiving a significant boost in various geoscientific domains (Brodrick et al., 2019; Hulbert et al., 2019; Rasp et al., 2018; Shen, 2018). Using AI techniques in geosciences has a long history. For example, the domain of hydroinformatics, formulated by Abbott (1991) 30 years ago, is defined as a union of computational hydraulics and AI (Solomatine & Ostfeld, 2008). However, the mainstream geoscientific community remains careful in adopting AI approaches, largely because of the assumed black-box nature of an AI model: it offers few mechanistic explanations beyond its fitting capability (Ebert-Uphoff et al., 2019;

Sun & Scanlon, 2019). Although some scientists have attempted to explain black-box models (Samek & Müller, 2019), Rudin (2019) recognizes that explaining black-box models rather than creating interpretable models in the first place is likely to perpetuate bad practice because the intentional explanations may be unreliable and even misleading. In addition to the reluctance in the community, data limitation is another obstacle (Karpatne et al., 2019).

Considering the respective merits of physical approaches and AI models, the synergy of the two paradigms has recently been envisioned as an attractive research topic in the geoscience community (Gil et al., 2019). Reichstein et al. (2019) summarized feasible strategies of physics-AI synergy in geoscience, some of which have been actively explored, such as using AI to build a surrogate model (Dwelle et al., 2019; Mo et al., 2019) or to learn and correct the mismatch between physical models and observations (Solomatine & Shrestha, 2009; Sun et al., 2019). In comparison, the strategy of really hybrid modeling, which focuses on a more physically realistic neural network (NN) through adding one or several physical layers, appears to be much less investigated (Reichstein et al., 2019). The hybrid modeling approach more closely meets the prospect of improving AI systems' geoscientific awareness, since the entire sequence of tasks is undertaken with a single and unified AI architecture (Figure S1 in the supporting information). Some studies in geoscience have attempted this strategy by introducing physical constraints into the loss function in the DL as a penalty term (Karpatne et al., 2017; Zhao et al., 2019) or wrapping analytical solutions of physical representations in DL as NN layers (de Bezenac et al., 2019; Wang et al., 2020). The existing approaches require physical knowledge to be expressed in a closed form. As the Earth is a highly dynamic system (Ghil, 2019), physical principles in geoscience are more often represented by differential equations (DEs) that cannot be analytically solved to achieve a closed-form solution. Consequently, how to design a network architecture for implanting geosystem dynamics when analytical solutions are not available remains an open question.

This study proposes an applicable solution to this problem by innovating a new NN architecture named physical process-wrapped recurrent neural network (denoted as P-RNN), which is designed to incorporate non-analytically solvable ordinary differential equations (ODEs) into DL models. ODEs are extensively employed by geoscientists for studying geosystem dynamics such as seismic signal analysis (Peng et al., 2014) and global climate modeling (Kaper & Engler, 2013) since the nineteenth century (Ghil, 2019). In this study, the new architecture was tested in the field of hydrology, using runoff modeling across the conterminous United States (CONUS) as an illustrative case. Hydrology is a cornerstone discipline in the geosciences (Fatichi et al., 2016), and catchment runoff modeling is the core task of hydrology (Beven, 2012). A catchment is a typical dynamic geosystem where water input (e.g., rainfall), output (e.g., runoff), and storage (e.g., soil moisture and snowpack) evolve over time. In this illustrative case, a conceptual hydrologic model (EXP-HYDRO; Patil & Stieglitz, 2014) is included as the P-RNN layer in a DL architecture, leading to hydrology-aware DL models. Enhanced simulation accuracy, robust cross-catchment transferability, and good intelligence for inferring unobserved processes were observed for the hydrology-aware DL models, highlighting the symbiotic integration between DL and the physical approach: the data-driven components fill gaps in the physical knowledge underlying the hydrologic model and the knowledge imparts physical awareness to the DL model. Overall, this study demonstrates that AI can garner physical knowledge the way humans do, if taught appropriately, and a symbiotic integration between DL and physical approaches is feasible, beneficial, and promising in advancing geoscience.

2. Methods

2.1. Generic Architecture

A physics-aware DL architecture comprising two data pipelines, each with several NN layers, is proposed (see Figure 1a). In the main pipeline, between the input (i.e., dynamic forcings) and output (i.e., dynamic responses), a P-RNN layer (to be elaborated in section 2.2) is constructed to wrap the geoscientific model (block a). The outputs of the P-RNN layer enter a series of common NN layers (block b), which can be fully connected layers, convolution layers, or recurrent layers (Goodfellow et al., 2016). In this design, the P-RNN layer endows the network with physical interpretability and consistency, and the common NN layers address processes unrepresented by the P-RNN layer (when the output of the P-RNN layer is different from the final output) or correct mismatch(when the output of the P-RNN layer is the same as the final output).

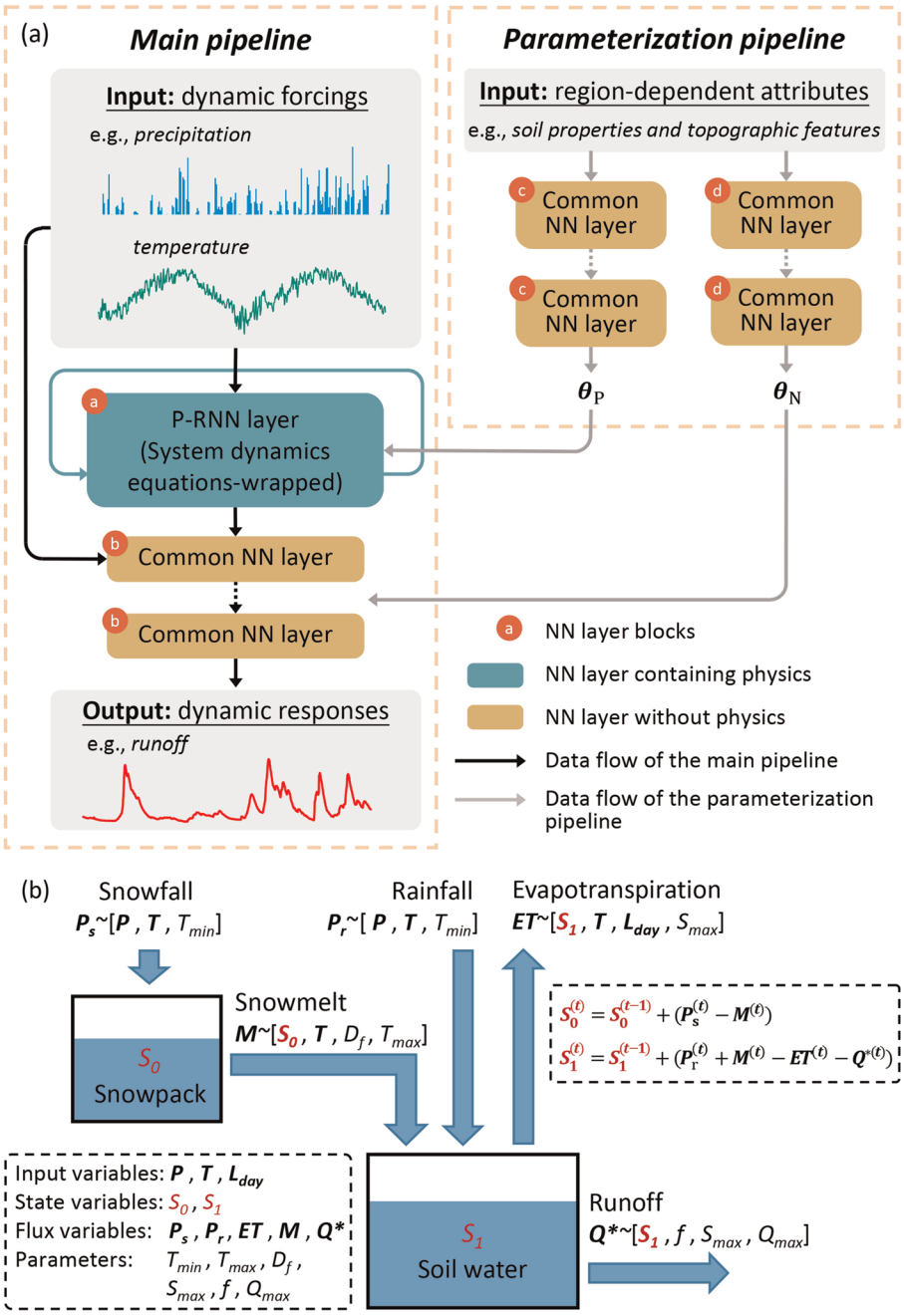

Figure 1. Embedding geosystem dynamics into deep learning architectures. (a) The proposed generic architecture that explicitly encodes geosystem dynamic behaviors. (b) Schematic diagram of the EXP-HYDRO model. The variables presented in the left dashed box are explained in Text S1. The state-space representations in the right dashed box are the step function of the physical process-wrapped recurrent neural network (P-RNN) layer in the case study.

The main pipeline contains two types of parameters: the architecture parameters θ_N (i.e., weights and biases) of the common NN layers and physically meaningful parameters θ_P of the P-RNN layer. To improve the transferability of the physics-aware DL across different regions, the architecture involves a parameterization pipeline for mapping region-dependent attributes (e.g., soil properties and topography) onto θ_N and θ_P using additional networks (blocks c and d). Once the architecture parameters of the parametrization pipeline itself are obtained, θ_N and θ_P can be determined on the interior. The unique design of a separate parameterization pipeline makes the architecture flexible to produce either global/regional models (the pipeline enabled) or

local models (the pipeline disabled) and ensures that the architecture can infer spatially varying θ_N and θ_P when addressing cross-region learning.

2.2. The P-RNN Layer

The proposed P-RNN layer is designed to encode geosystem dynamics as an NN architecture. A dynamic geosystem is often characterized by the state-space representation, which describes the system response for certain inputs and can be modeled by a combination of ODEs (state equations) and output equations in the following form:

$$\begin{cases} \frac{\mathrm{d}}{\mathrm{d}t}s(t) = F\left(s(t), x(t); \theta^{\langle F\rangle}\right) \\ y(t) = G\left(s(t), x(t); \theta^{\langle G\rangle}\right) \end{cases}. \tag{1}$$

where $s(t) \in \mathbb{R}^n$, $x(t) \in \mathbb{R}^p$, and $y(t) \in \mathbb{R}^q$, respectively, denote state vectors, input vectors, and output vectors with respect to time t, $F: \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^n$, $G: \mathbb{R}^n \times \mathbb{R}^p \to \mathbb{R}^q$, and $\theta^{\langle F\rangle}$ and $\theta^{\langle G\rangle}$ are the parameter sets.

Analogous to the ordinary recurrent neural network (RNN) architecture (Rumelhart et al., 1986), the backbone of the P-RNN layer is formed by recurrent cells that can provide memories of the past sequence (see Figure S2). Within recurrent cells in the P-RNN architecture, the connections between neurons (inputs x, states s, and outputs y) are specified with the state-space representation (Equation 1) in an explicit discrete form and the architecture parameters (i.e., weights and biases) of an ordinary RNN are replaced by parameters with physical meanings (i.e., θ_P). Table S1 in the supporting information summarizes the pseudocode of the P-RNN layer implemented in this study. Niu et al. (2019) demonstrated the connections between network architectures of the RNN family and numerical methods of ODEs, which theoretically supports the use of the P-RNN to tackle problems involving system dynamics.

2.3. Implementation for Runoff Modeling

In the case study, the EXP-HYDRO model (Patil & Stieglitz, 2014) is wrapped with the P-RNN layer. It is a conceptual, spatially lumped hydrologic model underpinned by two discrete state-space representations (Figure 1b). Although the model is not process-based (or physically based) per the rigorous definition in Fatichi et al. (2016), it adheres strictly to the law of mass conservation and has been demonstrated competent in large-sample catchments (Patil & Stieglitz, 2014). The snow accumulation bucket (S_0) and the storage in the catchment bucket (S_1) are the model state variables. Daily precipitation P, daily temperature T, and day length L_{day} are the model inputs, and the daily streamflow at the catchment outlet Q is the output of interest. The model involves six physically meaningful parameters (i.e., θ_P) that control the hydrologic behaviors (see Figure 1b). Detailed information on the model is provided in Text S1 in the supporting information.

In this case, the two pipelines in the generic architecture (Figure 1a) are respectively for runoff modeling and its parameterization. The inputs of the main pipeline are the meteorological forcing variables including P, T, and L_{day}, as required by the P-RNN layer, as well as the shortwave downward radiation *SRad* and vapor pressure *VP*. The EXP-HYDRO model-wrapped P-RNN layer provides a preliminary runoff estimation Q^*, which is fed into a two-layer common NN block together with the five input variables (see Figure S3). In this study, the 1-D convolution layers (Conv1D; Fukushima & Miyake, 1982) are utilized in the common NN block for considering the lagged influence of predictors in 10 days on current hydrologic responses (see Text S2 for details). The Conv1D layer is capable of handling the lagged effect by a one-direction convolution operation and has been employed for data-based hydrologic modeling in several studies (Feng et al., 2019). Through the Conv1D layers, the approximation errors associated with the physical approach are corrected, and the final runoff Q is derived. The model following the main pipeline scheme that hybridizes physical approaches (represented by the P-RNN layer) and data-driven components (i.e., Conv1D layers) is hereafter referred to as the "hybrid DL model." Furthermore, the parameterization pipeline is used to provide the catchment awareness for the main pipeline, in which two blocks of fully connected layers offer θ_P for the P-RNN layer and θ_N for the Conv1D layers. As a result of the parameterization pipeline, θ_N and θ_P can vary among different catchments with the physiographic attributes. More information regarding the network structures (e.g., number of layers and neurons) and training hyperparameters (e.g., training epoch and learning rate) of the model is elaborated in Text S2.

The data used in this study were obtained from the freely available CAMELS dataset (Addor et al., 2017; Newman et al., 2015), which consists of the area-averaged daily hydrometeorological time series and catchment attributes for 671 catchments across the CONUS with limited human disturbance. In this study, 569 catchments where daily hydrologic observations are complete and continuously recorded from 1 October 1980 to 30 September 2010 were considered (see Figure S4).

3. Results and Discussion

3.1. Synergy of Physics and AI

This study first built separate hybrid DL models for individual catchments, in which only the main pipeline in Figure 1a was activated, as each model only involves a single catchment where the spatial heterogeneity of parameters is beyond consideration. For each individual catchment, the hybrid DL model was compared against three other models. The first two models are variants (two partial models). The first variant is referred to as the "physical NN model," in which the main pipeline contains the "physical" part only (i.e., the P-RNN layer); the second variant is referred to as the "common NN model," in which the main pipeline only contains the "data-driven" part (i.e., the Conv1D layers). The third model is the long short-term memory (LSTM) network hydrologic model developed by Kratzert et al. (2018) for reference purposes. Both the common NN model and the LSTM model are purely data-driven, whereas the physical NN model is essentially a numeric variant of the EXP-HYDRO model. The four models were separately calibrated at 569 individual catchments for daily streamflow from 1 October 1980 to 30 September 2000.

Figure 2a presents the spatial distributions of the Nash-Sutcliffe efficiency (NSE) values, which measure the overall simulation performance (Nash & Sutcliffe, 1970), from the four models during the independent evaluation period (1 October 2000 to 30 September 2010). Visually, the hybrid DL model outperforms the other three in terms of the NSE in general. For all the models, the catchments with low NSE values are mainly located in the central arid regions of the United States in which runoff is largely controlled by short-duration, high-intensity precipitation events and infiltration-excess overland flow dominates runoff generation (Berghuijs et al., 2016). The prototypical model of the P-RNN layer operates with a mechanism of saturation-excess overland flow and is mainly applicable in the Northeast and the Pacific Northwest as well as forested mountain areas in the United States, which is indicated by the worst performance of the physical NN model outside these regions. It is also difficult for purely data-driven models such as the common NN model and the LSTM model to satisfactorily reproduce flashy hydrographs in such regions due to the infrequent storms and flood records. Such a spatial pattern of model performance was also revealed in previous studies (Essou et al., 2016; Herman et al., 2013; Kratzert et al., 2018; Newman et al., 2017; Sun et al., 2014) with similar explanations (McCabe & Wolock, 2011).

We further compared the models by scrutinizing the hydrographs they reproduce for the individual catchment. The hybrid DL model shows superior performance in reproducing flow peaks and capturing the overall pattern. The two purely data-driven models show unfavorable performance for flow peaks, and the physical NN model demonstrates mediocre overall performance. Figure S5 shows a detailed comparison of the hydrographs predicted by the four models at a typical catchment. Similar results were observed for the majority of other catchments, as indicated by Figure 2b, which illustrates the distributions of three goodness-of-fit metrics, including the NSE, the absolute value of peak flow bias (PFAB; Yilmaz et al., 2008), and mNSE (emphasizing the performance for baseflow; Legates & McCabe, 1999) evaluated for the four models. Table S2 compares the NSE performance of the four AI models in this study with those of previous models (Kratzert et al., 2018; Newman et al., 2017; Patil & Stieglitz, 2014; Xia et al., 2012), most of which are based on the same dataset and have an overlapping evaluation period with that in this study. Among the models compared, the hybrid DL model exhibits the highest overall performance.

Table S3 explores the correlation between the hybrid DL model and its two partial models in terms of the NSE performance within the 569 catchments. NSE $\geq$ 0.55 was considered the threshold for good performance (Knoben et al., 2019; Newman et al., 2015). If at least one of the partial models achieves good performance (see rows 1 to 3 and 5 to 7 of the table), the hybrid DL model performs satisfactorily in most cases, which implies that the predictive power of the partial models can be maintained by the hybrid model. More interestingly, even when both partial models do not meet the performance criterion, there is still a large chance (see row 4 of the table) that the hybrid model reaches a satisfactory NSE, which reminds us

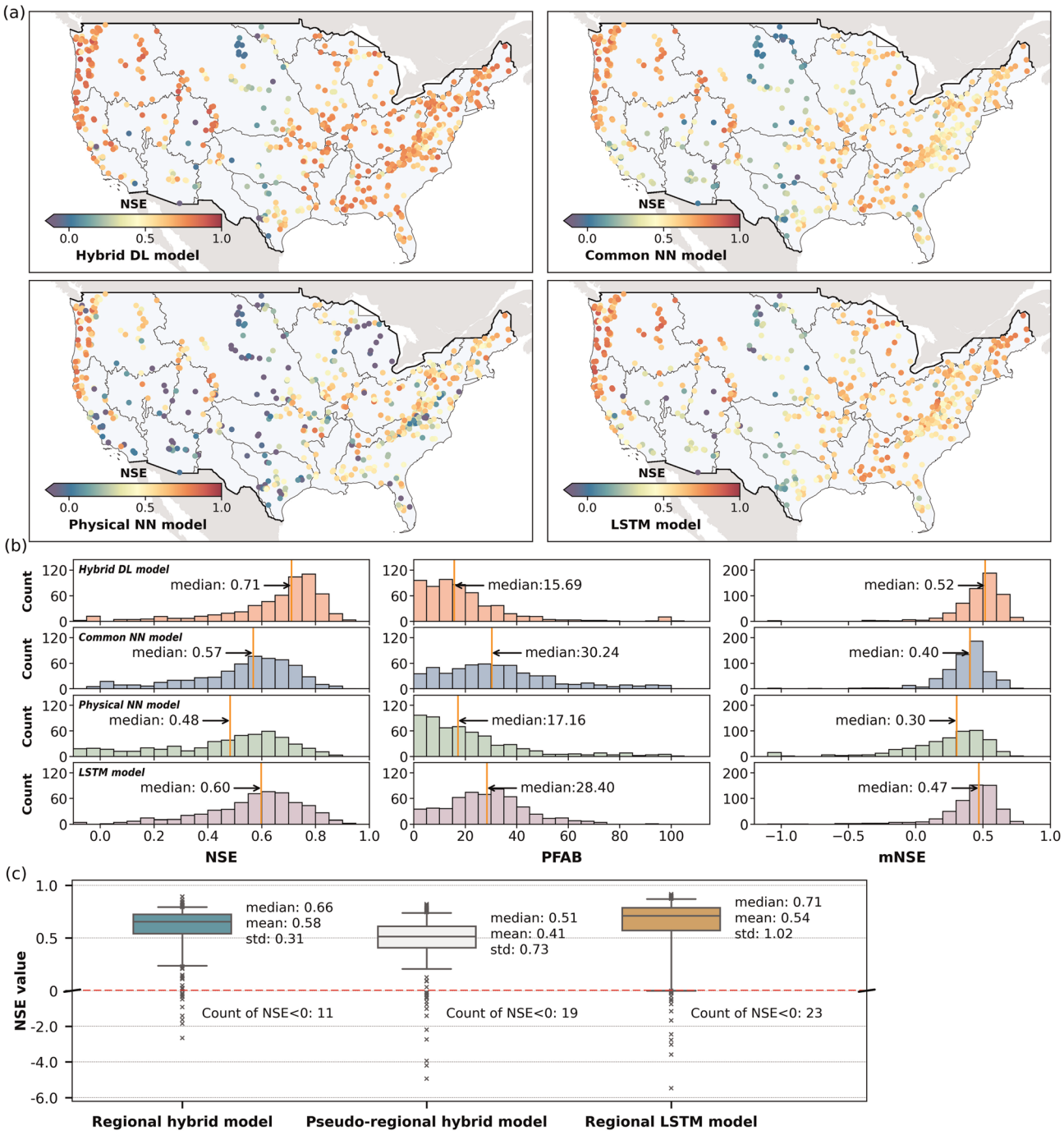

Figure 2. Performance and transferability of different artificial intelligence models. (a) Spatial distributions of the Nash-Sutcliffe efficiency (NSE) for the four individual models. The NSE colormap is capped within [0, 1] for better visualization. (b) Histograms of three metrics for the four models. (c) The transferability of the three regional models in 450 catchments for the fivefold cross validation. The box shows the interquartile range of the data, and the whiskers indicate the 5th and 95th percentiles. The *y* axis is capped within [−6, 1] in different scales for better visualization.

of the old saying "two heads are better than one." Further investigation reveals that the hybrid model can adequately maintain water balance when the prototypical model encoded in the P-RNN layer has a reliable performance in the target catchment (see Figure S6).

3.2. Robust Transferability of the Educated AI

Hydrologic predictions in ungauged basins are a great challenge in hydrology (Hrachowitz et al., 2013) and require better comprehension of the links between the hydrological function and physical properties of a catchment (Wagener et al., 2007). A typical way to address predictions in ungauged basins is to build a regional hydrologic model with site-specific parameterization (Oudin et al., 2008; Razavi & Coulibaly, 2013). In this study, a regional DL model is built following the new DL architecture that contains both the main pipeline and parameterization pipeline and is hereafter referred to as the "regional hybrid model."

As mentioned above, runoff modeling based on the CAMELS dataset is generally unsuccessful in the arid central part of the United States. Figure S7 suggests the large possibility that the hybrid DL model performs satisfactorily for an aridity index $\leq$1.2 or the frequency of dry days $\leq$280 days/year. Thus, 450 catchments (see Figure S8) were retained to evaluate the model's transferability via fivefold cross validation. Specifically, the 450 catchments were randomly and equally sampled into five groups. In each of the five runs, the regional hybrid model, which employs 27 available physiographic attributes (summarized in Table S4) as the inputs for parameterization, was trained during 1 October 1980 to 30 September 2010 in four groups of catchments (as gauged) and validated on the rest of the groups (as ungauged) during the same period. For comparison, two additional regional models were built and trained in the same way. One is the same as the regional hybrid model except that the parameterization pipeline is not activated, and therefore, θ_P and θ_N are constant across catchments. This model is hereafter referred to as the "pseudo-regional hybrid model." The other is the regional LSTM model by Kratzert et al. (2019), which uses the direct concatenation of catchment attributes and meteorological variables as inputs. The pseudo-regional hybrid model ignores the catchment attributes, and the regional hybrid model and regional LSTM model represent two distinct strategies of representing catchment attributes.

Figure 2c illustrates the box plots of the NSE values achieved by the three regional models in the cross validations. Compared with the NSE values from the pseudo-regional hybrid model, incorporating catchment attributes in the regional hybrid model significantly improves the performance, which confirms that the variations in physiographic attributes between catchments are critical to distinguishing hydrologic behaviors. Although the regional LSTM model exhibits slightly better performance than the regional hybrid model with respect to the median NSE, its predictive ability has significant fluctuations and the model showed catastrophically bad predictions (NSE < 0) in 23 catchments. The mean NSE achieved by the regional hybrid model is higher, the standard deviation is considerably lower (an important indication of a robust model), and fewer catastrophic predictions occur (see Figure S9), suggesting a much more robust transfer-learning ability of the regional hybrid model.

LSTM models are currently the popular machine learning approach for runoff modeling. The regional LSTM model in this study infuses the catchment attributes and meteorological forcing into one neuron perceptron (these values are directly weighted and summed) and achieved a slightly higher median NSE than the regional hybrid model, whereas the interpretable structure of the latter enhances the confidence of predictions in unfamiliar regions. As noted by Hrachowitz et al. (2013), linking catchment form (i.e., the attributes) to hydrological function (i.e., the way a catchment responds to input, including model parameters) is important for adequately addressing the challenge of ungauged basins. From this perspective, it is worth giving priority to the model's inherent interpretability even if it allows for some decrease in accuracy. Moreover, by dissecting the parameterization pipeline in a calibrated model, one can further explore spatial patterns of the inferred parameters, which offer insights into catchment processes and performance of the hydrologic model (e.g., Beck et al., 2013; Foks et al., 2019; Santhi et al., 2008). In this sense, the regional hybrid model can also be viewed as a streamlined parameterization module for the original hydrologic model, which is highly efficient in cross-region applications. Please refer to Text S3 and Figure S10 for more details.

3.3. Can the Educated AI Reason?

The previous sections show that the symbiotic integration between DL and physical approaches can realize better and more robust predictive power. However, the predictive power mainly reflects the induction capability of the educated AI rather than its deduction capability. The hydrologic model learned by the AI system describes runoff generation as well as the dynamics of snow accumulation and melting, and the P-RNN layer can output the water storage in snowpack (denoted as S_0) as an intermediate state variable. Therefore,

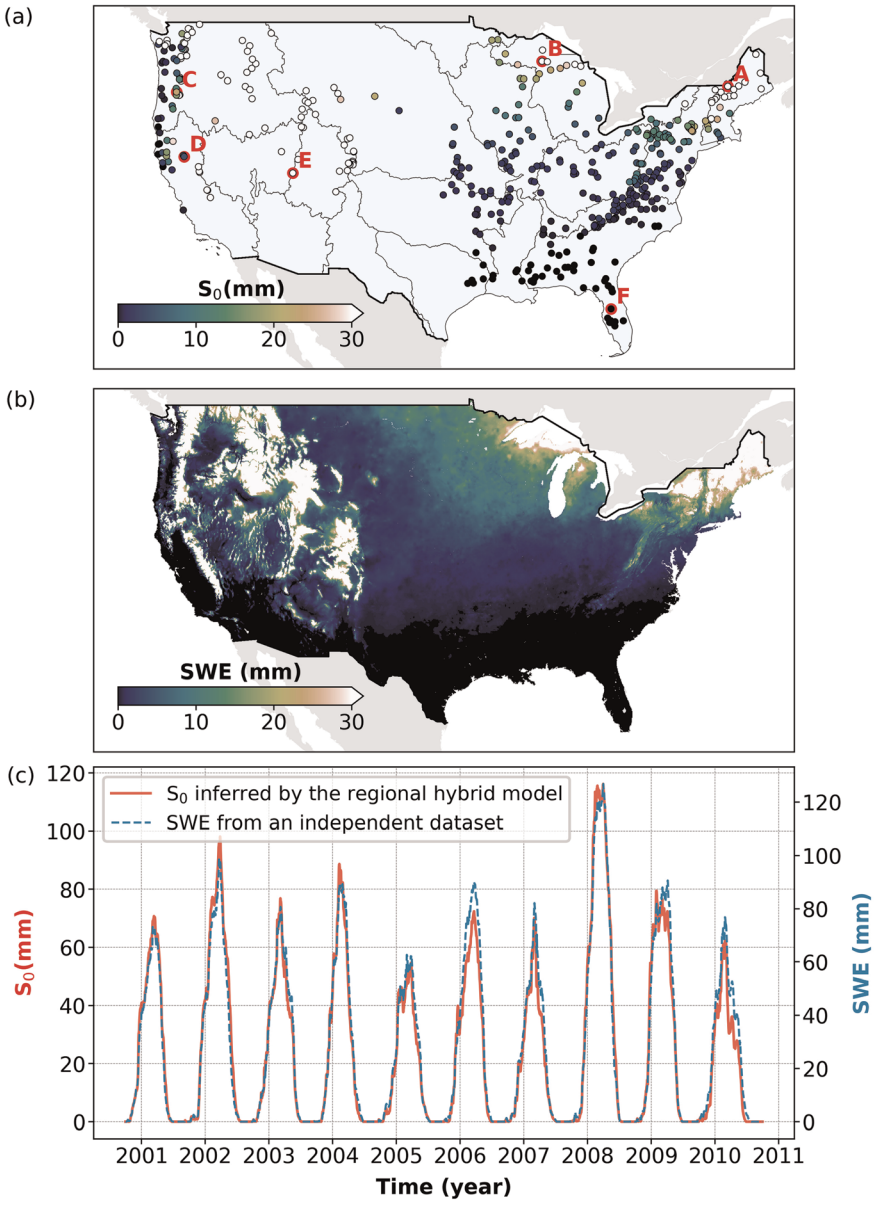

Figure 3. Comparison of the water storage in snowpack (S_0) inferred by the regional hybrid model with the snow water equivalent (SWE) data. (a) Ten-year average of daily S_0 in 450 catchments obtained from the P-RNN layer in the regional hybrid model. The highlighted catchments with letters were selected for the analyses in Figure S11. (b) Raster map of the 10-year average of daily SWE, based on the data from the NASA National Snow and Ice Data Center. (c) The time series of S_0 and SWE averaged over 450 catchments.

it is of great interest to determine whether the AI system trained with flow observations only can appropriately infer snow pattern and dynamics. An affirmative answer would serve as evidence for the reasoning capability of educated AI.

Figure 3a illustrates the average S_0 from 1 October 2000 to 30 September 2010 in all 450 catchments based on the regional hybrid model simulations. Intuitively and without exception, the catchments with noticeable snow accumulation have high altitudes or high latitudes. To verify the unsupervised learned S_0, daily snow water equivalent (SWE) data over the CONUS (4 km × 4 km spatial resolution) were obtained from the NASA National Snow and Ice Data Center (Zeng et al., 2018). Figure 3b presents the spatial distribution of the average SWE in the same period, and the spatial pattern coincides with that in Figure 3a. More

surprisingly, as Figure 3c shows, the snow dynamics inferred by the educated AI perfectly match those of the independent SWE dataset (with a correlation coefficient of 0.99). Similar comparisons were also made for individual catchments. Figure S11 demonstrates an excellent match for representative catchments (highlighted in Figure 3a). These promising results confirm that the DL model with physical awareness is capable of not merely undertaking the intentional task of data fitting, but more importantly, reasonable deduction of processes without direct observations.

3.4. Toward Smarter AI for Geoscience

This study seamlessly merged physical knowledge into DL architecture by creating a P-RNN layer to solve ODEs in discrete form, a common way to describe geosystem dynamics. It is important to note that the architecture presented is not necessarily an unchangeable template, considering the flexibility and easy extensibility of DL (e.g., modular building blocks). For example, the common NN layers noted in Figure 1a can be any type, such as the advanced ConvLSTM layers (Shi et al., 2015) that capture spatiotemporal correlations. Purely data-driven NN layers can also be put on top of the physics-aware NN layers to provide inputs (e.g., parameterizations for clouds (Rasp et al., 2018) and the fluid motion field (de Bezenac et al., 2019)).

In addition to solving ODEs with NN layers to reproduce 1D system dynamics, encoding partial differential equation (PDE) solvers in DL architectures to simulate higher-dimensional system dynamics deserves further exploration in geoscience. In recent years, DL has gradually emerged as a powerful technique for solving PDEs in the field of applied mathematics (Sirignano & Spiliopoulos, 2018). For example, the utilization of NN layer architectures involving iterative convolution operations is a promising solution for the combination of DL and higher-dimensional system dynamics (Long et al., 2018).

With regard to future applications in runoff modeling, one of the next steps would be to encode additional runoff generation mechanisms and hydrologic processes (e.g., subsurface flow) in the P-RNN layer, such that a higher transferability of regional models can be achieved.

4. Conclusion

This study represents a firm step toward realizing the vision of Reichstein et al. (2019) of tackling Earth system challenges using hybrid physics-AI approaches. We demonstrate that symbiotic integration is critical to the success of such approaches, as the data-driven component fills gaps in physical knowledge (e.g., processes unrepresented by physical approaches and physical parameters that are difficult to measure) and the physical component endows the AI system with physical consistency and reasoning ability. As the illustrative case of runoff modeling shows, the DL architecture proposed in this study enables an AI system to directly learn elements of physical representations and make accurate deductions on unobserved phenomena. In this regard, the novel framework presents a promising strategy to effectively educate AI with available geoscientific knowledge. We envision a number of future studies to implement the developed framework in various geoscientific contexts and further expand it to enable the use of smarter AI systems to advance research in the geoscience field.

Acknowledgments

This work was financially supported by the Strategic Priority Research Program of Chinese Academy of Sciences (XDA20100104) and the National Natural Science Foundation of China (51961125203, 91647201, and 41622111). We gratefully acknowledge the use of the dataset provided by (Addor et al., 2017; Newman et al., 2015) and the codes of the EXP-HYDRO model and LSTM model available in Kratzert et al. (2018), Kratzert et al. (2019), and Patil and Stieglitz (2014). The code of the models developed in this study is available online (https://doi.org/10.5281/zenodo.3856486).

References

Abbott, M. B. (1991). *Hydroinformatics: Information technology and the aquatic environment*. Aldershot: Avebury Technical.

Addor, N., Newman, A. J., Mizukami, N., & Clark, M. P. (2017). The CAMELS data set: Catchment attributes and meteorology for large-sample studies. *Hydrology and Earth System Sciences*, *21*(10), 5293–5313. https://doi.org/10.5194/hess-21-5293-2017

Bauer, P., Thorpe, A., & Brunet, G. (2015). The quiet revolution of numerical weather prediction. *Nature*, *525*(7567), 47–55. https://doi.org/10.1038/nature14956

Beck, H. E., van Dijk, A., Miralles, D. G., de Jeu, R. A. M., Bruijnzeel, L. A., McVicar, T. R., & Schellekens, J. (2013). Global patterns in base flow index and recession based on streamflow observations from 3394 catchments. *Water Resources Research*, *49*, 7843–7863. https://doi.org/10.1002/2013WR013918

Berghuijs, W. R., Woods, R. A., Hutton, C. J., & Sivapalan, M. (2016). Dominant flood generating mechanisms across the United States. *Geophysical Research Letters*, *43*, 4382–4390. https://doi.org/10.1002/2016GL068070

Beven, K. (2012). *Rainfall-runoff modelling: The primer*, (2nd ed.). Chichester, UK: John Wiley & Sons, Ltd. https://doi.org/10.1002/9781119951001

de Bezenac, E., Pajot, A., & Gallinari, P. (2019). Deep learning for physical processes: Incorporating prior scientific knowledge. *Journal of Statistical Mechanics: Theory and Experiment*, *2019*(12), 124009. https://doi.org/10.1088/1742-5468/ab3195

Brodrick, P. G., Davies, A. B., & Asner, G. P. (2019). Uncovering ecological patterns with convolutional neural networks. *Trends in Ecology & Evolution*, *34*(8), 734–745. https://doi.org/10.1016/j.tree.2019.03.006

Dwelle, M. C., Kim, J., Sargsyan, K., & Ivanov, V. Y. (2019). Streamflow, stomata, and soil pits: Sources of inference for complex models with fast, robust uncertainty quantification. *Advances in Water Resources*, *125*, 13–31. https://doi.org/10.1016/j.advwatres.2019.01.002

Ebert-Uphoff, I., Samarasinghe, S., & Barnes, E. (2019). Thoughtfully using artificial intelligence in Earth science. *Eos*, *100*. https://doi.org/10.1029/2019eo135235

Essou, G. R. C., Arsenault, R., & Brissette, F. P. (2016). Comparison of climate datasets for lumped hydrological modeling over the continental United States. *Journal of Hydrology*, *537*, 334–345. https://doi.org/10.1016/j.jhydrol.2016.03.063

Eyring, V., Cox, P. M., Flato, G. M., Gleckler, P. J., Abramowitz, G., Caldwell, P., et al. (2019). Taking climate model evaluation to the next level. *Nature Climate Change*, *9*(2), 102–110. https://doi.org/10.1038/s41558-018-0355-y

Fatichi, S., Vivoni, E. R., Ogden, F. L., Ivanov, V. Y., Mirus, B., Gochis, D., et al. (2016). An overview of current applications, challenges, and future trends in distributed process-based models in hydrology. *Journal of Hydrology*, *537*, 45–60. https://doi.org/10.1016/j.jhydrol.2016.03.026

Feng, D., Fang, K., & Shen, C. (2019). Enhancing streamflow forecast and extracting insights using long-short term memory networks with data integration at continental scales. *arXiv e-prints*, arXiv:1912.08949. Retrieved from https://arxiv.org/abs/1912.08949

Foks, S. S., Raffensperger, J. P., Penn, C. A., & Driscoll, J. M. (2019). Estimation of base flow by optimal hydrograph separation for the conterminous United States and implications for national-extent hydrologic models. *Water*, *11*(8), 1629. https://doi.org/10.3390/w11081629

Fukushima, K., & Miyake, S. (1982). Neocognitron: A new algorithm for pattern recognition tolerant of deformations and shifts in position. *Pattern Recognition*, *15*(6), 455–469. https://doi.org/10.1016/0031-3203(82)90024-3

Ghil, M. (2019). A century of nonlinearity in the geosciences. *Earth and Space Science*, *6*, 1007–1042. https://doi.org/10.1029/2019ea000599

Gil, Y., Pierce, S. A., Babaie, H., Banerjee, A., Borne, K., Bust, G., et al. (2019). Intelligent systems for geosciences: An essential research agenda. *Communications of the ACM*, *62*(1), 76–84. https://doi.org/10.1145/3192335

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. London: MIT Press. Retrieved from https://mitpress.mit.edu/books/deep-learning

Herman, J. D., Reed, P. M., & Wagener, T. (2013). Time-varying sensitivity analysis clarifies the effects of watershed model formulation on model behavior. *Water Resources Research*, *49*, 1400–1414. https://doi.org/10.1002/wrcr.20124

Hrachowitz, M., Savenije, H. H. G., Blöschl, G., McDonnell, J. J., Sivapalan, M., Pomeroy, J. W., et al. (2013). A decade of predictions in ungauged basins (PUB)—A review. *Hydrological Sciences Journal-Journal Des Sciences Hydrologiques*, *58*(6), 1198–1255. https://doi.org/10.1080/02626667.2013.803183

Hulbert, C., Rouet-Leduc, B., Johnson, P. A., Ren, C. X., Rivière, J., Bolton, D. C., & Marone, C. (2019). Similarity of fast and slow earthquakes illuminated by machine learning. *Nature Geoscience*, *12*(1), 69–74. https://doi.org/10.1038/s41561-018-0272-8

Kaper, H., & Engler, H. (2013). Chapter 12: Zonal energy budget. In *Mathematics and Climate*, (pp. 141–157). Philadelphia: Society for Industrial and Applied Mathematics. https://doi.org/10.1137/1.9781611972610.ch12

Karpatne, A., Atluri, G., Faghmous, J. H., Steinbach, M., Banerjee, A., Ganguly, A., et al. (2017). Theory-guided data science: A new paradigm for scientific discovery from data. *IEEE Transactions on Knowledge and Data Engineering*, *29*(10), 2318–2331. https://doi.org/10.1109/tkde.2017.2720168

Karpatne, A., Ebert-Uphoff, I., Ravela, S., Babaie, H. A., & Kumar, V. (2019). Machine learning for the geosciences: Challenges and opportunities. *IEEE Transactions on Knowledge and Data Engineering*, *31*(8), 1544–1554. https://doi.org/10.1109/tkde.2018.2861006

Knoben, W. J. M., Freer, J. E., & Woods, R. A. (2019). Technical note: Inherent benchmark or not? Comparing Nash-Sutcliffe and Kling-Gupta efficiency scores. *Hydrology and Earth System Sciences*, *23*(10), 4323–4331. https://doi.org/10.5194/hess-23-4323-2019

Kratzert, F., Klotz, D., Brenner, C., Schulz, K., & Herrnegger, M. (2018). Rainfall-runoff modelling using long short-term memory (LSTM) networks. *Hydrology and Earth System Sciences*, *22*(11), 6,005–6,022. https://doi.org/10.5194/hess-22-6005-2018

Kratzert, F., Klotz, D., Herrnegger, M., Sampson, A. K., Hochreiter, S., & Nearing, G. S. (2019). Toward improved predictions in ungauged basins: Exploiting the power of machine learning. *Water Resources Research*, *55*, 11,344–11,354. https://doi.org/10.1029/2019wr026065

Legates, D. R., & McCabe, G. J. (1999). Evaluating the use of "goodness-of-fit" measures in hydrologic and hydroclimatic model validation. *Water Resources Research*, *35*(1), 233–241. https://doi.org/10.1029/1998wr900018

Long, Y., She, X., & Mukhopadhyay, S. (2018). HybridNet: Integrating model-based and data-driven learning to predict evolution of dynamical systems. *arXiv e-prints*, arXiv:1806.07439. Retrieved from https://arxiv.org/abs/1806.07439

McCabe, G. J., & Wolock, D. M. (2011). Independent effects of temperature and precipitation on modeled runoff in the conterminous United States. *Water Resources Research*, *47*, W11522. https://doi.org/10.1029/2011wr010630

Mo, S. X., Zhu, Y. H., Zabaras, N., Shi, X., & Wu, J. (2019). Deep convolutional encoder-decoder networks for uncertainty quantification of dynamic multiphase flow in heterogeneous media. *Water Resources Research*, *55*, 703–728. https://doi.org/10.1029/2018wr023528

Nash, J. E., & Sutcliffe, J. V. (1970). River flow forecasting through conceptual models. Part I—A discussion of principles. *Journal of Hydrology*, *10*(3), 282–290. https://doi.org/10.1016/0022-1694(70)90255-6

Newman, A. J., Clark, M. P., Sampson, K., Wood, A., Hay, L. E., Bock, A., et al. (2015). Development of a large-sample watershed-scale hydrometeorological data set for the contiguous USA: Data set characteristics and assessment of regional variability in hydrologic model performance. *Hydrology and Earth System Sciences*, *19*(1), 209–223. https://doi.org/10.5194/hess-19-209-2015

Newman, A. J., Mizukami, N., Clark, M. P., Wood, A. W., Nijssen, B., & Nearing, G. (2017). Benchmarking of a physically based hydrologic model. *Journal of Hydrometeorology*, *18*(8), 2215–2225. https://doi.org/10.1175/jhm-d-16-0284.1

Niu, M., Horesh, L., & Chuang, I. (2019). Recurrent neural networks in the eye of differential equations. *arXiv e-prints*, arXiv:1904.12933. Retrieved from https://arxiv.org/abs/1904.12933

Oudin, L., Andreassian, V., Perrin, C., Michel, C., & Le Moine, N. (2008). Spatial proximity, physical similarity, regression and ungaged catchments: A comparison of regionalization approaches based on 913 French catchments. *Water Resources Research*, *44*, W03413. https://doi.org/10.1029/2007wr006240

Patil, S., & Stieglitz, M. (2014). Modelling daily streamflow at ungauged catchments: What information is necessary? *Hydrological Processes*, *28*(3), 1159–1169. https://doi.org/10.1002/hyp.9660

Peng, H., Kitagawa, G., Takanami, T., & Matsumoto, N. (2014). State-space modeling for seismic signal analysis. *Applied Mathematical Modelling*, *38*(2), 738–746. https://doi.org/10.1016/j.apm.2013.07.008

Rasp, S., Pritchard, M. S., & Gentine, P. (2018). Deep learning to represent subgrid processes in climate models. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(39), 9684–9689. https://doi.org/10.1073/pnas.1810286115

Razavi, T., & Coulibaly, P. (2013). Streamflow prediction in ungauged basins: Review of regionalization methods. *Journal of Hydrologic Engineering*, *18*(8), 958–975. https://doi.org/10.1061/(asce)he.1943-5584.0000690

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Rudin, C. (2019). Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. *Nature Machine Intelligence*, *1*(5), 206–215. https://doi.org/10.1038/s42256-019-0048-x

Rumelhart, D. E., Hinton, G. E., & Williams, R. J. (1986). Learning representations by back-propagating errors. *Nature*, *323*(6088), 533–536. https://doi.org/10.1038/323533a0

Samek, W., & Müller, K.-R. (2019). Towards explainable artificial intelligence. In *Explainable AI: Interpreting, explaining and visualizing deep learning*, (pp. 5–22). Cham, Switzerland: Springer Nature. https://doi.org/10.1007/978-3-030-28954-6_1

Santhi, C., Allen, P. M., Muttiah, R. S., Arnold, J. G., & Tuppad, P. (2008). Regional estimation of base flow for the conterminous United States by hydrologic landscape regions. *Journal of Hydrology*, *351*(1–2), 139–153. https://doi.org/10.1016/j.jhydrol.2007.12.018

Shen, C. P. (2018). A transdisciplinary review of deep learning research and its relevance for water resources scientists. *Water Resources Research*, *54*, 8558–8593. https://doi.org/10.1029/2018wr022643

Shi, X., Chen, Z., Wang, H., Yeung, D., Wong, W., Woo, W. (2015). Convolutional LSTM network: A machine learning approach for precipitation nowcasting. *arXiv e-prints*, arXiv:1506.04214. Retrieved from https://arxiv.org/abs/1506.04214

Sirignano, J., & Spiliopoulos, K. (2018). DGM: A deep learning algorithm for solving partial differential equations. *Journal of Computational Physics*, *375*, 1339–1364. https://doi.org/10.1016/j.jcp.2018.08.029

Solomatine, D. P., & Ostfeld, A. (2008). Data-driven modelling: Some past experiences and new approaches. *Journal of Hydroinformatics*, *10*(1), 3–22. https://doi.org/10.2166/hydro.2008.015

Solomatine, D. P., & Shrestha, D. L. (2009). A novel method to estimate model uncertainty using machine learning techniques. *Water Resources Research*, *45*, W00B11. https://doi.org/10.1029/2008WR006839

Sun, A. Y., & Scanlon, B. R. (2019). How can big data and machine learning benefit environment and water management: A survey of methods, applications, and future directions. *Environmental Research Letters*, *14*(7), 073001. https://doi.org/10.1088/1748-9326/ab1b7d

Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., & Zhong, Z. (2019). Combining physically based modeling and deep learning for fusing GRACE satellite data: Can we learn from mismatch? *Water Resources Research*, *55*, 1179–1195. https://doi.org/10.1029/2018WR023333

Sun, A. Y., Wang, D. B., & Xu, X. L. (2014). Monthly streamflow forecasting using Gaussian process regression. *Journal of Hydrology*, *511*, 72–81. https://doi.org/10.1016/j.jhydrol.2014.01.023

Wagener, T., Sivapalan, M., Troch, P., & Woods, R. (2007). Catchment classification and hydrologic similarity. *Geography Compass*, *1*(4), 901–931. https://doi.org/10.1111/j.1749-8198.2007.00039.x

Wang, N., Zhang, D., Chang, H., & Li, H. (2020). Deep learning of subsurface flow via theory-guided neural network. *Journal of Hydrology*, *584*, 124700. https://doi.org/10.1016/j.jhydrol.2020.124700

Xia, Y. L., Mitchell, K., Ek, M., Cosgrove, B., Sheffield, J., Luo, L., et al. (2012). Continental-scale water and energy flux analysis and validation for North American Land Data Assimilation System project phase 2 (NLDAS-2): 2. Validation of model-simulated streamflow. *Journal of Geophysical Research: Atmospheres*, *117*, D03110. https://doi.org/10.1029/2011JD016051

Yang, C., Huang, Q., Li, Z., Liu, K., & Hu, F. (2016). Big data and cloud computing: Innovation opportunities and challenges. *International Journal of Digital Earth*, *10*(1), 13–53. https://doi.org/10.1080/17538947.2016.1239771

Yilmaz, K. K., Gupta, H. V., & Wagener, T. (2008). A process-based diagnostic approach to model evaluation: Application to the NWS distributed hydrologic model. *Water Resources Research*, *44*, W09417. https://doi.org/10.1029/2007WR006716

Zeng, X., Broxton, P., & Dawson, N. (2018). Snowpack change from 1982 to 2016 over conterminous United States. *Geophysical Research Letters*, *45*, 12,940–12,947. https://doi.org/10.1029/2018GL079621

Zhao, W. L., Gentine, P., Reichstein, M., Zhang, Y., Zhou, S., Wen, Y., et al. (2019). Physics-constrained machine learning of evapotranspiration. *Geophysical Research Letters*, *46*, 14,496–14,507. https://doi.org/10.1029/2019GL085291