

# Understanding the Physics Representation of Deep Learning Models in Environmental Applications

September 2022

1 Xiaodong Chen  
2 Ruby Leung  
3 Sally Wang  
4 Zhuoran Duan

## DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency thereof, nor Battelle Memorial Institute, nor any of their employees, **makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights.** Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof, or Battelle Memorial Institute. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

PACIFIC NORTHWEST NATIONAL LABORATORY  
*operated by*  
BATTELLE  
*for the*  
UNITED STATES DEPARTMENT OF ENERGY  
*under Contract DE-AC05-76RL01830*

Printed in the United States of America

Available to DOE and DOE contractors from  
the Office of Scientific and Technical  
Information,  
P.O. Box 62, Oak Ridge, TN 37831-0062  
[www.osti.gov](http://www.osti.gov)  
ph: (865) 576-8401  
fox: (865) 576-5728  
email: [reports@osti.gov](mailto:reports@osti.gov)

Available to the public from the National Technical Information Service  
5301 Shawnee Rd., Alexandria, VA 22312  
ph: (800) 553-NTIS (6847)  
or (703) 605-6000  
email: [info@ntis.gov](mailto:info@ntis.gov)  
Online ordering: <http://www.ntis.gov>

# Understanding the Physics Representation of Deep Learning Models in Environmental Applications

September 2022

1 Xiaodong Chen  
2 Ruby Leung  
3 Sally Wang  
4 Zhuoran Duan

Prepared for  
the U.S. Department of Energy  
under Contract DE-AC05-76RL01830

Pacific Northwest National Laboratory  
Richland, Washington 99354

## Abstract

Deep learning (DL) models have been popular in earth and environmental modeling and analysis, which exhibit huge potential in capturing and reconstructing the non-linearity of relevant environmental processes. They are extensively used as analytical tools or emulators for multiple domains (atmosphere, land surface, ocean, and biogeochemistry). Despite their success, their internal working mechanism remains largely unknown. Such a lack of knowledge hinders the identification of physically consistent models that are fully adaptive to non-stationary climate, as well as the development of physics-informed machine learning such as physics-informed neural network (PINN). To establish preliminary knowledge and framework of such physics representation evaluation, this project focuses on an improved understanding of DL models in the environmental applications. DL models are increasingly applied to environmental modeling and prediction. However, they have been evaluated mostly from a performance perspective, and there is a gap in understanding how they represent the known physics internally. Such knowledge is especially critical when applying DL models under climate change conditions, where new inputs are likely outside the ranges of the training datasets.

In this project, we reveal how the known physical processes are represented within DL models from both statistical and mechanistic perspectives. Leveraging the traditional model evaluations that focus more on the accuracies of predictions, we establish a framework that examines both the accuracy and physics representation of DL models. This analysis framework can identify DL models that make the correct predictions based on correct physics, thus enhancing the existing explainable artificial intelligence (explainable-AI) portfolio. It lays a foundation for developing novel metrics to evaluate the emerging DL models in environmental applications. This knowledge also informs the development of physics-informed DL models by revealing the direct connections between the known physical processes and specific model components or structures.

## Acknowledgments

This research was supported by the Earth and Biological Sciences Investment, under the Directed Research and Development (LDRD) Program at Pacific Northwest National Laboratory (PNNL). The computational work was performed using PNNL Research Computing at Pacific Northwest National Laboratory. PNNL is a multi-program national laboratory operated for the U.S. Department of Energy (DOE) by Battelle Memorial Institute under Contract No. DE-AC05-76RL01830.

## Contents

Abstract.....	ii
Acknowledgments.....	iii
1.0 Introduction .....	1
2.0 Data.....	3
3.0 Model development and evaluation .....	7
3.1 Model development.....	7
3.2 Model evaluation .....	9
3.3 Model verification .....	11
4.0 References.....	13

## Figures

Figure 1.	Understanding the physics presentation of deep learning models can help achieve better integration into earth system modeling. ....	2
Figure 2.	Overview of SNOTEL snow measurements. Left: 829 active SNOTELs across the western United States and Alaska. Right: daily time series of precipitation, SWE, and air temperature data for one SNOTEL as an example. Source: <a href="https://www.pnnl.gov/data-products">https://www.pnnl.gov/data-products</a> . ....	3
Figure 3.	Correlation analysis of the SWE and SNOTEL-derived monthly meteorological statistics in the same month. Higher correlations (towards 1 or -1) indicate a stronger enhancement or inhibition effects between these two factors, respectively. The values in the diagonal grids are the self-correlations which are all equal to 1. ....	5
Figure 4.	Correlation analysis of the SWE (of current month) and SNOTEL-derived monthly meteorological statistics in the previous month. This figure is similar to Figure 3 but with meteorological statistics taken from the previous month rather than the current month. ....	5
Figure 5.	Sensitivity of snowpack response as a function of mean air temperature.....	6
Figure 6.	Performance of a DNN model (mode #1). Left: the evolution of loss as a function of epochs during model training. Blue line indicates the loss over the training subset, and orange line indicates the loss over validation subset. Smaller loss indicates better model performance. Right: Evaluation of the predicted monthly SWE against the observed SWE. Overall this DNN achieved an $R^2$ of 0.901. ....	7
Figure 7.	Structures of two LSTM models. Left (model #2): the model that uses the meteorological condition of previous 12 months as input for the prediction of SWE at the current month. Right (model #3): the model that uses the meteorological condition of previous 6 months as input for the prediction of SWE at the current month. ....	8
Figure 8.	Performance of a LSTM model (model #2). Left: the evolution of model performance during model training. Right: Evaluation of the predicted	

monthly SWE against the observed SWE. Overall this LSTM achieved an  $R^2$  or 0.91. .... 8

Figure 9. Performance of a LSTM model (model #3). Left: the evolution of loss as a function of epochs during model training. Right: Evaluation of the predicted monthly SWE against the observed SWE. Overall this DNN achieved an  $R^2$  or 0.91. .... 9

Figure 10. Statistical evaluation of DL models. Here the SHAP score of model #1 is shown, and model inputs are ranked by their |SHAP| values on the y-axis. Higher |SHAP| values indicate more important features. .... 10

Figure 11. Physics evaluation of DL models. Here the sensitivity of models #2 and #3 to mean monthly temperature are shown, along with the observations (green boxes). .... 11

## Tables

Table 1. SNOTELmonthly statistics derived from the daily data. .... 3

Table 2. DL model performance under different climates..... 11

## 1.0 Introduction

Deep learning (DL), an emerging branch of machine learning (ML), has been widely explored and increasingly adopted in environmental research in recent years (Shen 2018). They are extensively used as analytical tools or emulators for multiple domains (atmosphere, land surface, ocean, and biogeochemistry) (Shi 2020; Anderson and Lucas 2018; Sahoo et al. 2019; L. Wang, Li, and Lv 2022; Hsieh 2020; He et al. 2016).

DL/ML models can represent complex non-linear relationships between predictors and predictands, making them powerful tools for simulating the dynamics of environmental systems across a wide range of spatial-temporal scales. Therefore, they have been extensively tested as alternatives to physics models at module or model levels (Fang and Shen 2020; Kratzert et al. 2018). The Department of Energy has also prioritized ML and artificial intelligence (AI) development for predictive modeling and simulation across the physical sciences. Despite successful demonstrations, explaining the internal mechanisms of DL/ML models remains a challenge, with only preliminary progress achieved so far (Kratzert et al. 2019; S. S.-C. Wang et al. 2021). Most of these efforts are from a mathematical perspective, with domain knowledge seldom incorporated in these interpretations.

Meanwhile, recent research has focused on synergizing ML/DL models with established physics knowledge as part of the explainable-AI efforts (Xie et al. 2021; Kratzert et al. 2018; Fang and Shen 2020). These efforts generally fall into two categories: 1) configuring ML/DL models with structures that explicitly reflect known physics, such as mass/energy conservations or storage-flux relationship (physics-informed ML/DL); and 2) adding established physics knowledge as constraints to model structure or approximations (physics-constrained ML/DL). However, most of these efforts are still exploratory, and an improved, physics-based understanding of ML/DL is required to provide explicit guidance on model developments. Such an improved understanding also ensures that the established ML/DL models represent physics correctly, so they are adaptive to the changing environment. Through the proposed work, we will carry out a preliminary yet systematic evaluation of the physics representation in the DL models. The knowledge obtained from this project will enhance the existing explainable-AI portfolio and guide the development of physics-informed ML models.

This project is both an implementation and enhancement of the existing explainable-AI techniques. The obtained knowledge about DL models in this project will benefit the relevant environmental analysis and earth system model (e.g., E3SM) development activities, which implement AI/DL to describe various dynamics of environmental processes (Figure 1).

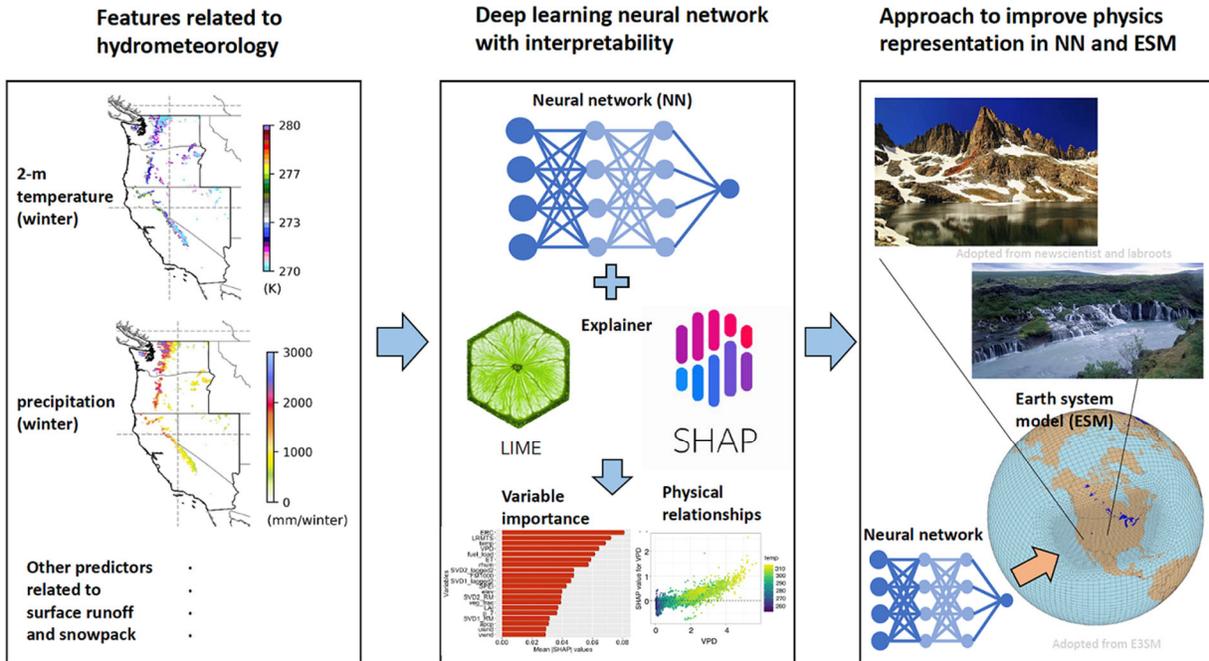


Figure 1. Understanding the physics presentation of deep learning models can help achieve better integration into earth system modeling.

## 2.0 Data

Snow Telemetry (SNOTEL) ground measure of snowpack condition (snow water equivalent, or SWE) to develop DL models (Yan et al. 2018). The raw data is obtained as the quality-controlled version developed at PNNL (<https://www.pnnl.gov/data-products>). They are distributed across the western US, and records from 800+ sites are used in our work (Figure 2). Their records cover 1981-2020, with records length varying across different sites. These data are also used in various regional hydrological and hydroclimate model evaluations (Chen et al. 2019).

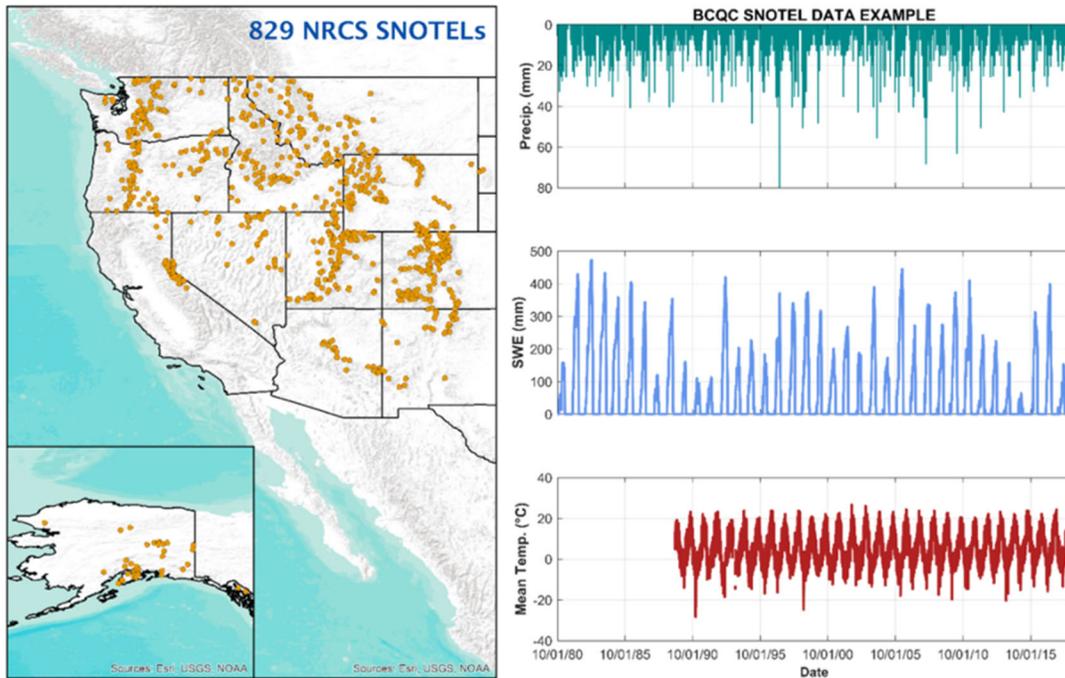


Figure 2. Overview of SNOTEL snow measurements. Left: 829 active SNOTELs across the western United States and Alaska. Right: daily time series of precipitation, SWE, and air temperature data for one SNOTEL as an example. Source: <https://www.pnnl.gov/data-products>.

Based on the daily observed temperature records, a total of 20 statistics at the monthly step are derived, as summarized in Table 1.

Table 1. SNOTEL monthly statistics derived from the daily data.

Index	Variable	Description
1	Fwet	Fraction of wet days (i.e., precipitating days)
2	Ptot	Total precipitation in this month
3	Pstd	Standard deviation of daily precipitation
4	Tmean-all	Mean temperature of all days
5	Tmean-std	Standard deviation of daily mean temperature
6	Tmean-wet-std	Tmean-std over precipitating days only

7	Tmean-Tp	Precipitation-weighted temperature, kind of the average temperature of precipitation
8	Tmean-dry-std	Tmean-std over non-precipitating days only
9	Tmax-wet-max	Maximum of daily max temperature over wet days
10	Tmax-wet-mean	Mean of daily max temperature over wet days
11	Tmax-wet-min	Minimum of daily max temperature over wet days
12	Tmax-dry-max	(similar but for non-precipitating days)
13	Tmax-dry-mean	
14	Tmax-dry-min	
15	Tmin-wet-max	(similar but for daily minimum temperature)
16	Tmin-wet-mean	
17	Tmin-wet-min	
18	Tmin-dry-max	
19	Tmin-dry-mean	
20	Tmin-dry-min	
21	SWE-1	The snow condition at the beginning (i.e., 1 <sup>st</sup> day) of this month. Predictand of DL models

(a) wet days are defined as days with precipitation amount higher than 0.1 inch

Figure 3 shows the correlation of these inputs within the same month, which highlights a strong correlation between the SWE on the 1<sup>st</sup> day of this month (SWE\_1) and the mean SWE of the same month (SWE\_mean). This indicates a strong tendency of SWE time series, which is physically reasonable as snowpack, especially in winter, is an accumulated response of snow accumulation/melting over multiple months. Regarding the meteorological conditions, temperature-relevant metrics show stronger correlations, indicating the snowpack response is more dependent on temperature.

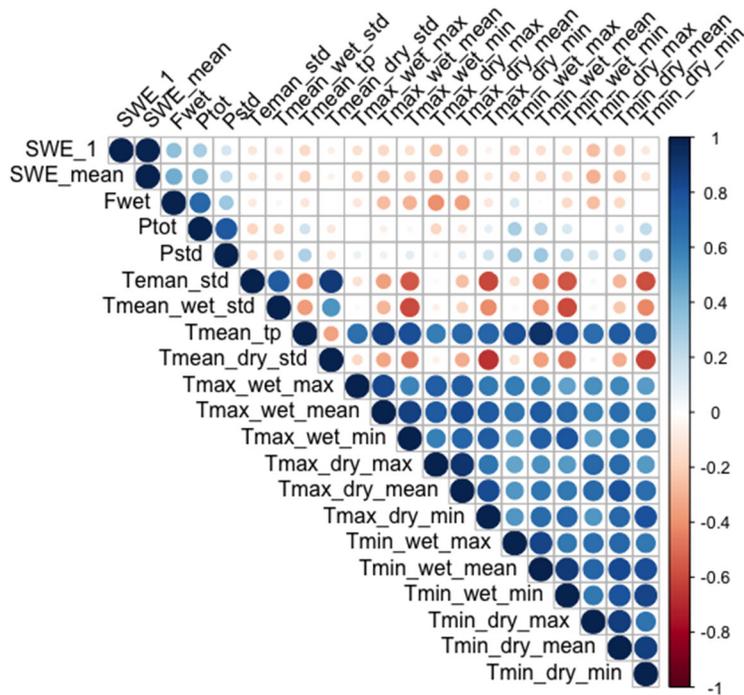


Figure 3. Correlation analysis of the SWE and SNOTEL-derived monthly meteorological statistics in the same month. Higher correlations (towards 1 or -1) indicate a stronger enhancement or inhibition effects between these two factors, respectively. The values in the diagonal grids are the self-correlations which are all equal to 1.8.

Figure 4 evaluates the SWE of the current month (SWE\_1 and SWE\_mean) with the meteorological and SWE (SWE\_1\_m1, SWE\_mean\_m1) of the previous month. The high positive connections between current SWE and SWE of the previous month (the upper right points) also confirm such lagged correlations.

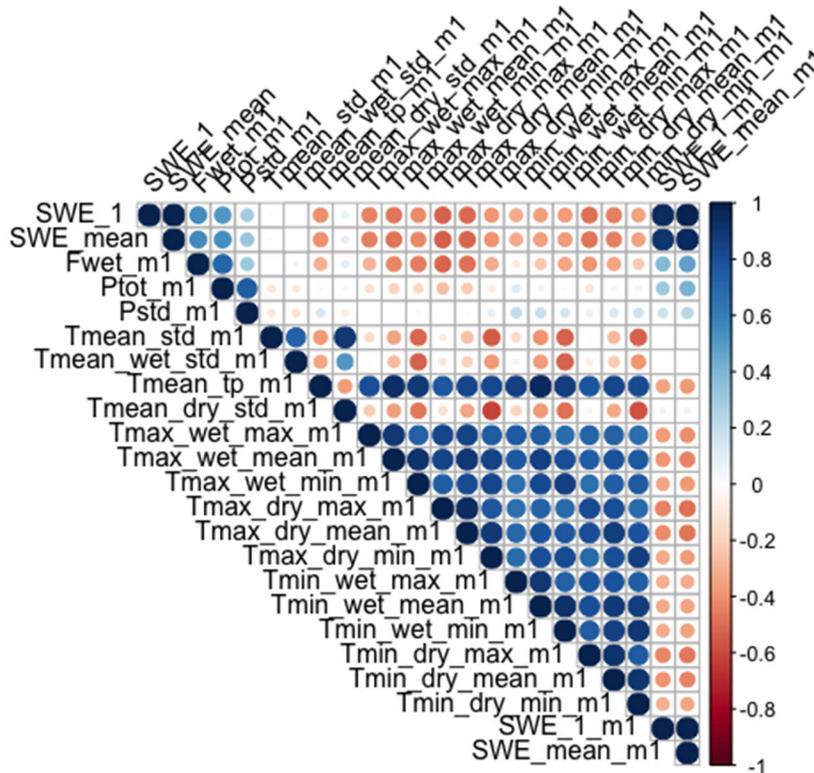


Figure 4. Correlation analysis of the SWE (of current month) and SNOTEL-derived monthly meteorological statistics in the previous month. This figure is similar to Figure 3 but with meteorological statistics taken from the previous month rather than the current month.

Given such a accumulated influence of meteorological conditions on monthly SWE, it makes more sense to evaluate the change of SWE in a specific month ( $\Delta SWE$ ) to meteorological conditions. Figure 5 illustrates the response of  $\Delta SWE/P$  to the mean temperature of the same month, and it is clear that snowpack accumulates when the mean temperature is lower than 0 °C, while intense melting occurs when temperature rises. This is consistent with our understanding, which also serves as the prior “physics knowledge” and is used to evaluate various DL models.

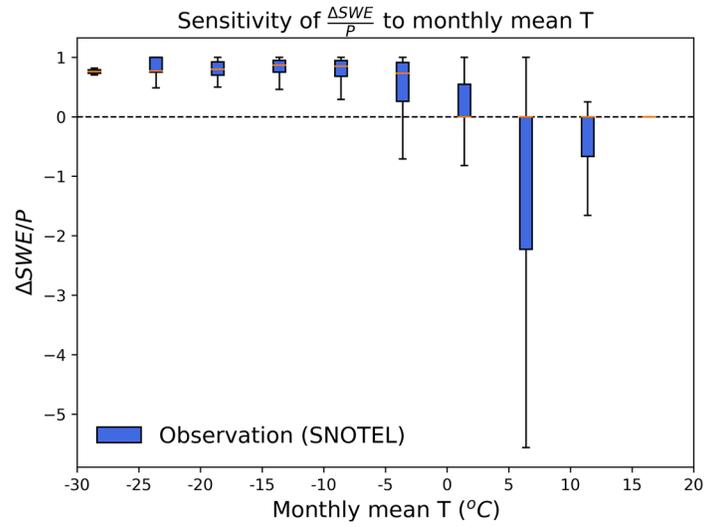


Figure 5. Sensitivity of snowpack response as a function of mean air temperature.

## 3.0 Model development and evaluation

### 3.1 Model development

Multiple models are developed for monthly snowpack condition (SWE) prediction. Specifically, we constructed dense neural network (DNN) and long short-term memory (LSTM) models to achieve good SWE prediction. We set a prior requirement of  $R^2 \geq 0.9$  to identify good models for the subsequent analyses.

Figure 6 illustrates the model development of one DNN model (model #1) that eventually achieved  $R^2=0.901$  in the SWE prediction. To predict the SWE of the current month, this model uses the full meteorological metrics along with the SWE condition in the previous month as inputs (Table 1). The network can be characterized as a 21x200x150x150x1 structure that contains three hidden layers. In total, we evaluated 6 DNN models, and this model #1 is the one with good performance throughout the analysis (including those analyses in section 3.2).

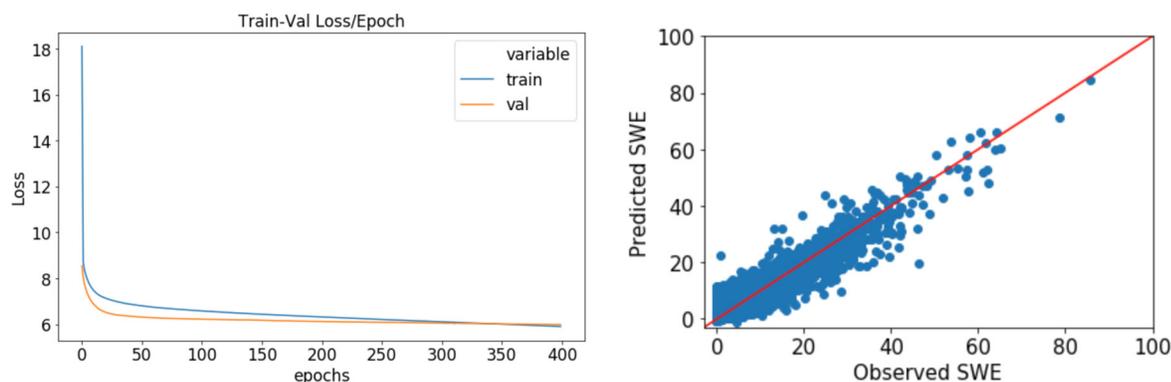


Figure 6. Performance of a DNN model (model #1). Left: the evolution of loss as a function of epochs during model training. Blue line indicates the loss over the training subset, and orange line indicates the loss over validation subset. Smaller loss indicates better model performance. Right: Evaluation of the predicted monthly SWE against the observed SWE. Overall this DNN achieved an  $R^2$  of 0.901.

As revealed in Figures 3 and 4, SWE has a strong lagged correlation (with SWE of previous months). This indicates the potential benefit of accounting for the meteorological conditions in the previous several months by the model. Therefore, LSTM models are employed to capture such long-term influence (Fang and Shen 2020; Xie et al. 2021; Jiang et al. 2022), with the structure of two good models illustrated in Figure 7. Both of these models use the following 6 meteorological metrics: Fwet, Ptot, Tmean-Tp, Tmean-all, Pstd, Tmean-std. So overall, model #2 uses 72 inputs (6 metrics x 12 months) to predict the current SWE, while model #3 uses 36 inputs for the same prediction. It is necessary to note that we identified a total of 5 LSTM models with  $R^2 > 0.9$ , each featuring metrics of 3, 6, 12, 18, and 24 months, respectively.

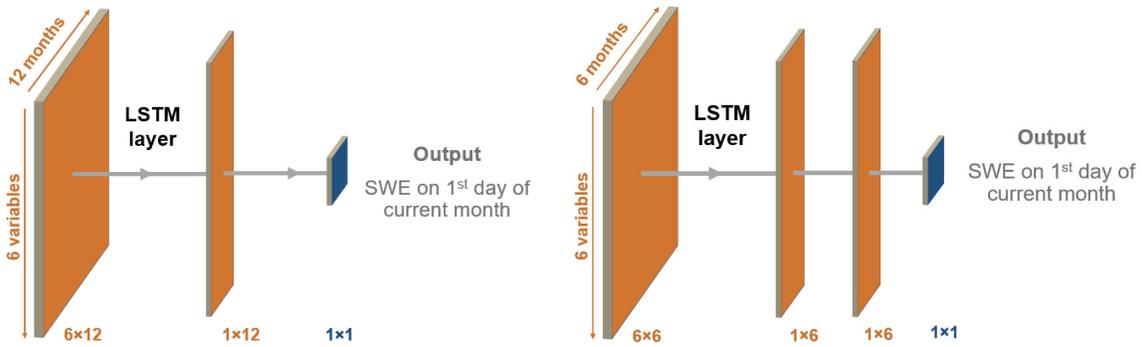


Figure 7. Structures of two LSTM models. Left (model #2): the model that uses the meteorological condition of previous 12 months as input for the prediction of SWE at the current month. Right (model #3): the model that uses the meteorological condition of previous 6 months as input for the prediction of SWE at the current month.

Figure 8 illustrates the model development of one LSTM model (model #2) that eventually achieved  $R^2=0.91$  in the SWE prediction.

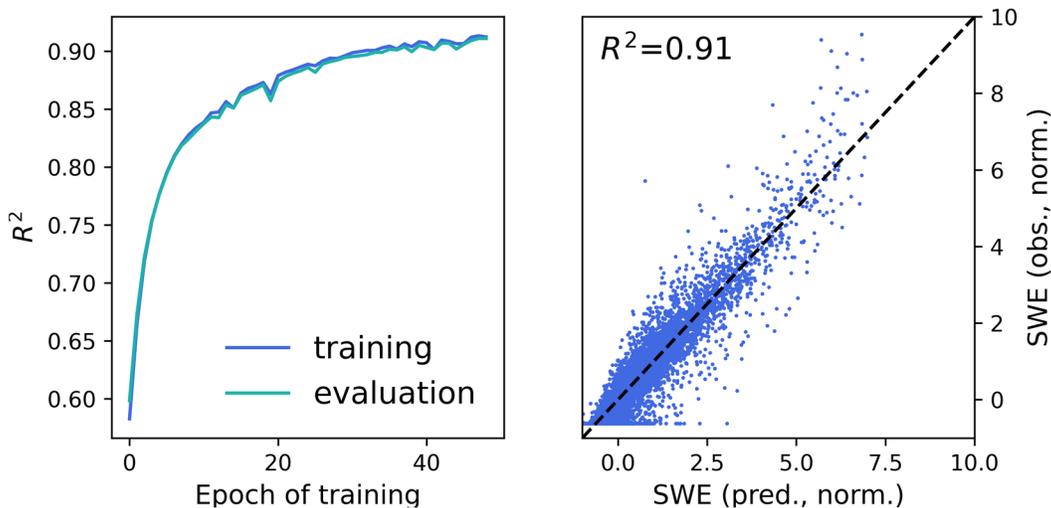


Figure 8. Performance of a LSTM model (model #2). Left: the evolution of model performance during model training. Right: Evaluation of the predicted monthly SWE against the observed SWE. Overall this LSTM achieved an  $R^2$  or 0.91.

Figure 9 illustrates another LSTM model (model #3) that achieved an overall accuracy of  $R^2=0.91$ . Compared to similar models but trained over the modeled data (i.e., the output of regional climate model output), its performance is reduced from  $R^2$  or  $\sim 0.96$  (Chen et al. 2021). This reflects the fact that SNOTEL data includes more randomness compared to model output.

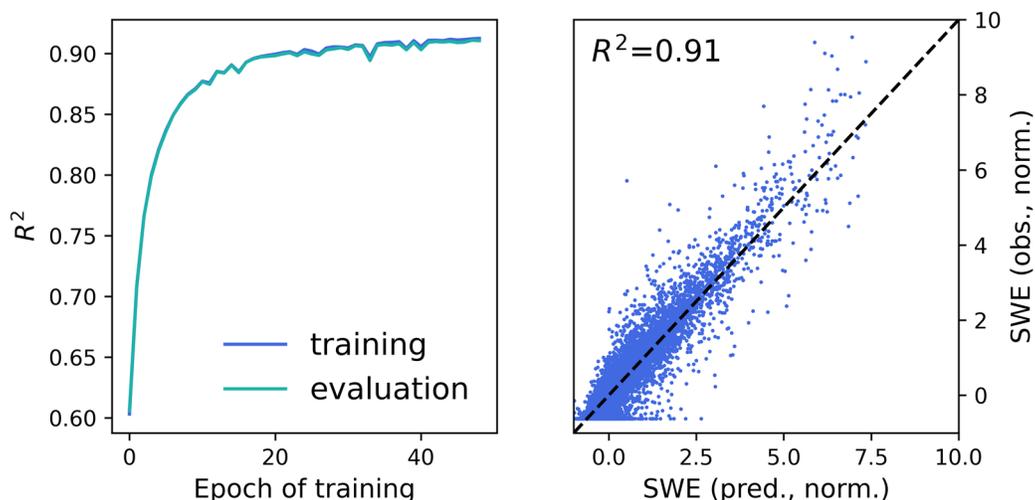


Figure 9. Performance of a LSTM model (model #3). Left: the evolution of loss as a function of epochs during model training. Right: Evaluation of the predicted monthly SWE against the observed SWE. Overall this DNN achieved an  $R^2$  or 0.91.

All these models exhibit prediction  $R^2 > 0.9$  so they are good models in the prediction perspective. They together form the model pool for the subsequent analysis, which will filter out those bad models from both statistical or physical perspectives.

### 3.2 Model evaluation

Shapley additive explanation (SHAP) is an emerging tool for machine learning model evaluation and interpretation, which has also been used in various studies (S. S.-C. Wang et al. 2021; Silva, Keller, and Hardin 2022). Its introduction can be found in (Štrumbelj and Kononenko 2014). In short, it tracks each input and calculates its contribution to the final prediction. Then a score can be assigned to each input (or feature), which allows cross-comparison of different features in terms of their relative importance. The higher the SHAP values, the more important a given feature is in determining the final prediction. From this perspective, SHAP evaluates the model “mechanism” in a qualitative perspective, and it is desired that good (or correct) models show a ranking of important features that are consistent with our established understanding of the same environmental process.

Figure 10 illustrates an example of SHAP analysis on DNN model (model #1). As it clearly shows, the snow conditions of the previous month have dominant impacts on the overall prediction, which is consistent with our knowledge that snowpack shows lagged correlation in time. Comparing temperature and precipitation metrics, Figure 10 indicates that temperature is generally more important than precipitation, which is also consistent with our analysis in Figures 3 and 4. Thus from this perspective, model #1 is a good model, which makes correct prediction by considering the true important inputs.

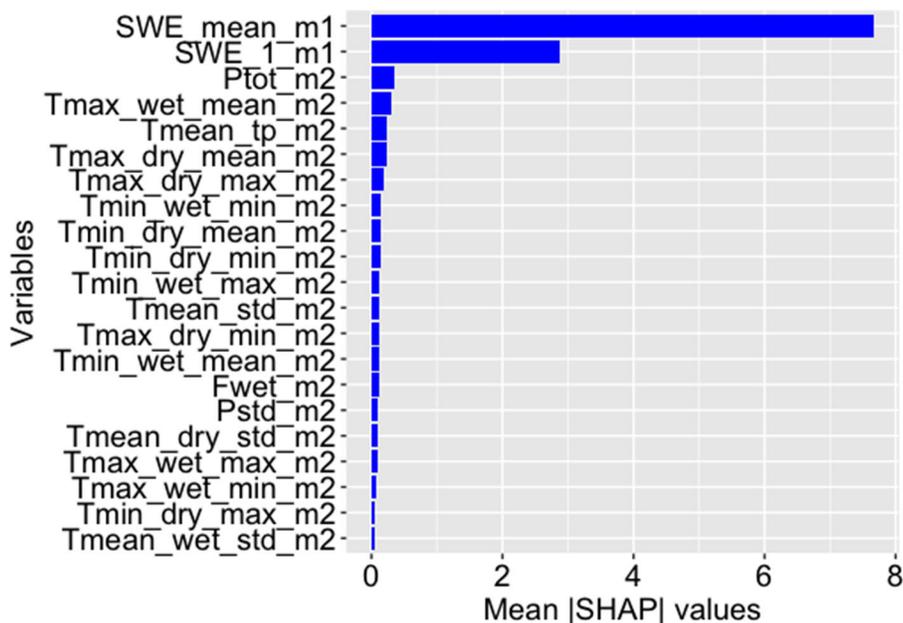


Figure 10. Statistical evaluation of DL models. Here the SHAP score of model #1 is shown, and model inputs are ranked by their |SHAP| values on the y-axis. Higher |SHAP| values indicate more important features.

SHAP describes which inputs are more important, yet how the models respond to these important inputs is to be examined. Sensitivity analysis, which relates the unit change of a given input to the corresponding change of model output, can be used to reveal such responses. The dependency of snowpack response to temperature (Figure 5) is an example of observed sensitivity, which can be used to evaluate the model sensitivity. By perturbing the monthly T of the final month (i.e., 1-month lead time input) and examining the change of SWE in the prediction, we can derive the monthly  $\Delta$ SWE to temperature in the DL models. They are illustrated in Figure 11, where both models #2 and #3 show similar responses (i.e., intense snowmelt when the temperature is above 0 °C). It is also notable that model #3 shows amplified sensitivities, where the magnitudes of pink boxes are larger than both observation and model #2. The implication of such amplified sensitivity will be discussed later. It is also important to note that both models #2 and #3 pass the SHAP analysis by showing physically reasonable important inputs.

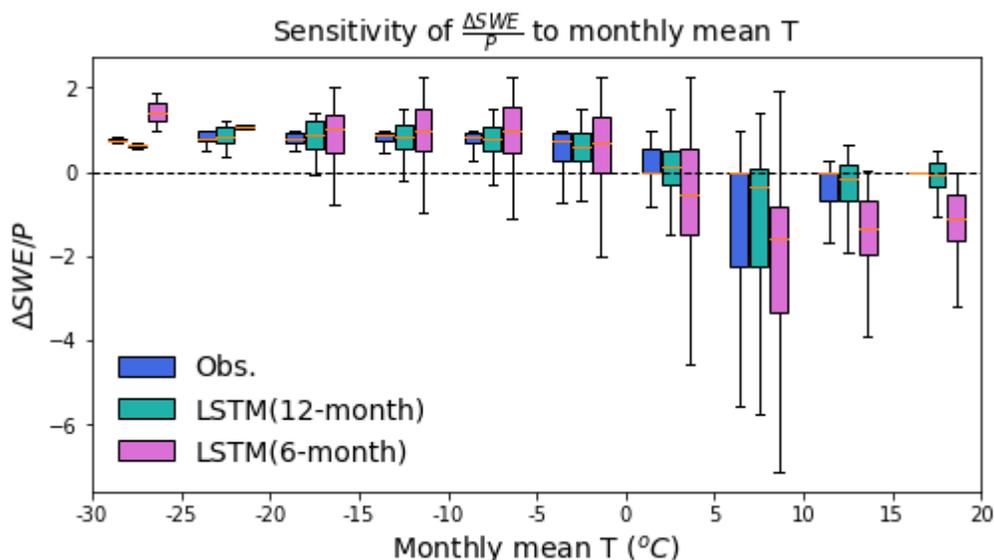


Figure 11. Physics evaluation of DL models. Here the sensitivity of models #2 and #3 to mean monthly temperature are shown, along with the observations (green boxes).

### 3.3 Model verification

The slightly different climate scenarios were used in this task. The 40-year SNOTEL data was split into two subsets: 1) 1981-2000; 2) 2001-2020. The first subset was used to re-train all the models described above, and the second subset was used to evaluate the model performance in a warmer climate. Should the model capture the physics of snowpack response correctly, such incremental extrapolation would exhibit similar performance as that over the training subset (i.e., the first subset).

Our analysis reveals varying performances of these models. Table 2 illustrates an interesting case using models #2 and #3. As reflected in Figures 8 and 9, they have similar performance over the SWE prediction, and this is also confirmed in the re-training of models using 1981-2000 data. However, applying them to a new climate, model #2 retains a similar performance as in the training period, and a slight reduction is consistent with prior expectations on such out-of-bag samples or extrapolation. However, the performance of model #3 deteriorates quickly in this new climate, with  $R^2$  quickly reduced to 0.76. This inexplicitly indicates that model #3 is not really presenting the physical connections between the meteorological conditions and snowpack response.

Table 2. DL model performance under different climates.

Model	$R^2$ (1981-2000)	$R^2$ (2001-2020)
LSTM (model #2): 12-month input	0.90	0.88
LSTM (model #3): 6-month input	0.90	0.76

To understand the behavior of model #3, we can examine Figure 11 again. Model #3 shows amplified sensitivity to temperature as compared to observation, which is likely caused by the short inputs used in this model: model #3 only reads 6-month meteorological data to make prediction. Since snowpack exhibits a strong seasonal cycle, 6-month is insufficient for snowpack to accumulate, especially in winter. Therefore, model #2 has to assume that more snowaccumulatin/melt occurs under the same temperature (as compared to model #2 which reads 12-month inputs) to make good predictions. Such inconsistency with observation means that the internal physics in model #3 is incorrect, and this issue gets highlighted in the new climate (2001-2020).

These results highlight the importance of correct representation of physics for those models that are adaptive to changing climate. Such representation cannot be revealed by SHAP or sensitivities alone: 1) SHAP alone only suggests which features are important, but how models respond to these features is still not evaluated (model #2 versus #3); 2) sensitivity analysis alone cannot ensure that all the non-important features are correctly “neglected” by the model. Suppose model X responds to temperature with similar sensitivity as observation, but it incorrectly responds to other trivial inputs with larger but canceling sensitivities. In this case, the model prediction may still look good (since the internal errors are canceled), but the model is still not presenting the correct physics. Therefore, a combination of SHAP with sensitivity results should be concurrently considered in DL model assessment, which would shape the new model evaluation metrics.

## 4.0 References

- Anderson, Gemma J., and Donald D. Lucas. 2018. "Machine Learning Predictions of a Multiresolution Climate Model Ensemble." *Geophysical Research Letters* 45 (9): 4273–80. <https://doi.org/10.1029/2018GL077049>.
- Chen, Xiaodong, L Ruby Leung, Yang Gao, and Ying Liu. 2021. "Response of U.S. West Coast Mountain Snowpack to Local Sea Surface Temperature Perturbations: Insights from Numerical Modeling and Machine Learning." *Journal of Hydrometeorology* 22 (4): 1045–62. <https://doi.org/10.1175/JHM-D-20-0127.1>.
- Chen, Xiaodong, L Ruby Leung, Mark Wigmosta, and Marshall Richmond. 2019. "Impact of Atmospheric Rivers on Surface Hydrological Processes in Western U.S. Watersheds." *Journal of Geophysical Research: Atmospheres* 124 (16): 8896–8916. <https://doi.org/10.1029/2019JD030468>.
- Fang, Kuai, and Chaopeng Shen. 2020. "Near-Real-Time Forecast of Satellite-Based Soil Moisture Using Long Short-Term Memory with an Adaptive Data Integration Kernel." *Journal of Hydrometeorology* 21 (3): 399–413. <https://doi.org/10.1175/JHM-D-19-0169.1>.
- He, Xiaogang, Nathaniel W Chaney, Marc Schleiss, and Justin Sheffield. 2016. "Spatial Downscaling of Precipitation Using Adaptable Random Forests." *Water Resources Research* 52 (10): 8217–37. <https://doi.org/10.1002/2016WR019034>.
- Hsieh, William W. 2020. "Improving Predictions by Nonlinear Regression Models from Outlying Input Data."
- Jiang, Shijie, Yi Zheng, Chao Wang, and Vladan Babovic. 2022. "Uncovering Flooding Mechanisms Across the Contiguous United States Through Interpretive Deep Learning on Representative Catchments." *Water Resources Research* 58 (1): e2021WR030185. <https://doi.org/10.1029/2021WR030185>.
- Kratzert, Frederik, Mathew Herrnegger, Daniel Klotz, Sepp Hochreiter, and Günter Klambauer. 2019. "NeuralHydrology – Interpreting LSTMs in Hydrology." In *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*, edited by Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller, 347–62. Cham: Springer International Publishing. [https://doi.org/10.1007/978-3-030-28954-6\\_19](https://doi.org/10.1007/978-3-030-28954-6_19).
- Kratzert, Frederik, D Klotz, C Brenner, K Schulz, and M Herrnegger. 2018. "Rainfall–Runoff Modelling Using Long Short-Term Memory (LSTM) Networks." *Hydrology and Earth System Sciences* 22 (11): 6005–22. <https://doi.org/10.5194/hess-22-6005-2018>.
- Sahoo, Bibhuti Bhusan, Ramakar Jha, Anshuman Singh, and Deepak Kumar. 2019. "Long Short-Term Memory (LSTM) Recurrent Neural Network for Low-Flow Hydrological Time Series Forecasting." *Acta Geophysica* 67 (5): 1471–81. <https://doi.org/10.1007/s11600-019-00330-1>.
- Shen, Chaopeng. 2018. "A Transdisciplinary Review of Deep Learning Research and Its Relevance for Water Resources Scientists." *Water Resources Research* 54 (11): 8558–93. <https://doi.org/10.1029/2018WR022643>.

- Shi, Xiaoming. 2020. "Enabling Smart Dynamical Downscaling of Extreme Precipitation Events With Machine Learning." *Geophysical Research Letters* 47 (19): e2020GL090309. <https://doi.org/https://doi.org/10.1029/2020GL090309>.
- Silva, Sam J, Christoph A Keller, and Joseph Hardin. 2022. "Using an Explainable Machine Learning Approach to Characterize Earth System Model Errors: Application of SHAP Analysis to Modeling Lightning Flash Occurrence." *Journal of Advances in Modeling Earth Systems* 14 (4): e2021MS002881. <https://doi.org/https://doi.org/10.1029/2021MS002881>.
- Štrumbelj, Erik, and Igor Kononenko. 2014. "Explaining Prediction Models and Individual Predictions with Feature Contributions." *Knowledge and Information Systems* 41 (3): 647–65. <https://doi.org/10.1007/s10115-013-0679-x>.
- Wang, Liwen, Qian Li, and Qi Lv. 2022. "Self-Supervised Classification of Weather Systems Based on Spatiotemporal Contrastive Learning." *Geophysical Research Letters* 49 (15): e2022GL099131. <https://doi.org/10.1029/2022GL099131>.
- Wang, Sally S.-C., Yun Qian, L Ruby Leung, and Yang Zhang. 2021. "Identifying Key Drivers of Wildfires in the Contiguous US Using Machine Learning and Game Theory Interpretation." *Earth's Future* 9 (6): e2020EF001910. <https://doi.org/10.1029/2020EF001910>.
- Xie, Kang, Pan Liu, Jianyun Zhang, Dongyang Han, Guoqing Wang, and Chaopeng Shen. 2021. "Physics-Guided Deep Learning for Rainfall-Runoff Modeling by Considering Extreme Events and Monotonic Relationships." *Journal of Hydrology* 603: 127043. <https://doi.org/10.1016/j.jhydrol.2021.127043>.
- Yan, Hongxiang, Ning Sun, Mark Wigmosta, Richard Skaggs, Zhangshuan Hou, and Ruby Leung. 2018. "Next-Generation Intensity-Duration-Frequency Curves for Hydrologic Design in Snow-Dominated Environments." *Water Resources Research* 54 (2): 1093–1108. <https://doi.org/10.1002/2017WR021290>.

# **Pacific Northwest National Laboratory**

902 Battelle Boulevard  
P.O. Box 999  
Richland, WA 99354

1-888-375-PNNL (7665)

***[www.pnnl.gov](http://www.pnnl.gov)***