ERDC TN-24-3
May 2024

Data-Driven Modeling of Groundwater Level Using Machine Learning

by Sourav Dutta, Anna M. Wagner, Theadora K. Hall, and Nawa Raj Pradhan

PURPOSE: This US Army Engineer Research and Development Center (ERDC) technical note documents a preliminary study on the use of specialized machine learning (ML) methods to model the variations in groundwater level (GWL) with time. This approach uses historical groundwater observation data at seven gage locations in Wyoming, USA, available from the USGS database and historical data on several relevant meteorological variables obtained from the ERA5 reanalysis dataset produced by the Copernicus Climate Change Service (usually referred to as C3S) at the European Center for Medium-Range Weather Forecasts to predict future GWL values for a desired period of time. The results presented in this report indicate that the ML method has the potential to predict both short-term (4-hourly) as well as daily variations in GWL several days into the future for the chosen study region, thus alleviating the need for employing sophisticated process-based numerical models with complicated model structure configurations.

INTRODUCTION: The availability, cost, and quality of water supplies are critical concerns for military installations and military operations within the United States and around the world. Military installations depend on reliable and secure sources of water to ensure mission execution, to both sustain humans and to support industrial and operational processes. In many regions, surface water resources are already fully allocated under nondrought conditions, and groundwater remains the main source for drought conditions. For a priority mission sustainment challenge, groundwater wells could be used as a primary water source. Knowing and understanding groundwater distribution in and around installations helps in identifying alternate sources of water like backup wells when the primary source of water is cut off or compromised (Smith et al. 2022). Hence, the estimation of GWL variations in time is critical, especially for any mission where the limited source of water is the groundwater. Unlike surface water and runoff, groundwater table modeling and analysis is computationally expensive and qualitatively uncertain (Chen et al. 2020). The primary reason is the difficulty in accessing the hydrogeological information (Bear et al. 1992), which requires a complicated model configuration. Therefore, it is important to develop alternative data-driven methods (e.g., ML-based) that require limited information and produce accurate results using limited computational resources. This study focuses on employing a supervised ML approach, which utilizes historical groundwater observations and globally available satellite-based hydrometeorological data to model and predict the future temporal trends of the water table fluctuations.

Traditional modeling methods for groundwater level prediction are complex and time consuming, so many ML-based methods have been developed for fast prediction of groundwater levels. Mahammad et al. (2023) compared the performance of the following six popular ML regression algorithms for this task: linear regression, decision tree regressor, support vector regressor, random forest regressor, K-nearest neighbor regressor, and extreme gradient boost regressor. This study

found that the extreme gradient boost performed best and the decision tree performed worst (Mahammad et al. 2023).

The rapid developments of deep neural network (DNN)–based methods in many areas of computational science have also led to their application in predicting GWLs. Recurrent neural networks (RNN) such as long-short-term memory networks (LSTM), convolutional neural networks (CNN), and genetic programming have been applied for this task (Cai et al. 2021). Most ML-based methods for predicting GWLs focus on a smaller region to improve the accuracy of the models. Cai et al. (2021) attempted to create a more general model for the eastern continental United States. This study used data from the Catchment Attributes and Meteorology for Large-Sample Studies (CAMELS) dataset provided by the National Center for Atmospheric Research for data on the precipitation, streamflow, GWL, daily mean temperature, day length, shortwave downward radiation, and vapor pressure at each of the catchments studied. Daily average GWLs were collected from the dataset provided by the USGS. A gated recurrent unit (GRU) was used to translate the surface water data from the CAMELS dataset into groundwater predictions with decent results for such a large area of study (Cai et al. 2021).

With the goal of using less input data to make ML methods for predicting GWLs more accessible to underdeveloped countries, Ao et al. (2021) compared the performance of Kernel-based Nonlinear Extension of the Arps (usually referred to as KNEA), decline model LSTM, and GRU on several combinations of input variables. The input variables included 1-month prior groundwater levels, air temperature, global solar radiation, precipitation, and amount of irrigation. For more generalizable models, LSTM performed the best. The study found that the most important input variables were temperature and 1-month prior groundwater levels. The amount of irrigation did not have much impact on the performance of the model (Ao et al. 2021).

METHOD: The emergence of modern ML methods has provided a transformative approach to effectively approximate and accelerate existing numerical models by leveraging the capabilities to incorporate multifidelity data streams from diverse sources, seamlessly explore massive design spaces, and identify complex, multivariate correlations. A variety of data-driven, ML-based approximation frameworks have been proposed to model the evolution of time-series data. Some of the highly successful examples involve the use of DNNs (Hesthaven et al. 2018), RNN (Wunsch et al. 2021), LSTM networks (Wan et al. 2018; Dutta et al. 2022), neural ordinary differential equations (Chen et al. 2018; Dutta et al., "Data-Driven Reduced Order Modeling," 2021; Dutta et al., "Neural Ordinary Differential Equations," 2021), and temporal convolutional networks (Wu et al. 2020).

An LSTM network is a special type of RNN that is well suited for performing classification and regression tasks based on time-series data. The main difference between the traditional RNN and the LSTM architecture is the addition of an LSTM memory cell that has the capability to retain information over time and an internal gating mechanism that regulates the flow of information in and out of the memory cell (Greff et al. 2016). The LSTM cell consists of three parts, also known as gates, that have specific functions. The first gate, called the *forget gate*, determines whether the information from the previous step in the sequence is to be remembered or can be forgotten. The second gate is called the *input gate*, which learns new information from the current input to this cell. The third and final gate, named the *output gate*, passes the updated information from the current step to the next step in the sequence. The basic LSTM equations for an arbitrary input vector u are the following:

- Input gate: $\zeta_i = \alpha_S \circ F_i(u)$
- Forget gate: $\zeta_f = \alpha_S \circ F_f(u)$
- Cell state: $c_t = \zeta_f \odot c_{t-1} + \zeta_i \odot (\alpha_T \circ F_a(u))$
- Output gate: $\zeta_o = \alpha_S \circ F_o(u)$
- Output: $h_t = \zeta_o \circ \alpha_T(c_t)$

where F refers to a linear transformation defined by a matrix multiplication and bias addition; that is $F(x) = Wx + b$, where $W \in R^{n \times m}$ is a matrix containing the layer weights, $b \in R^n$ is a vector containing bias values, and $x \in R^m$ is a vector containing the layer activations. Also, α_s and α_T denote the sigmoid and hyperbolic tangent activation functions, which are typically the default choices in an LSTM network, and $x \odot y$ denotes a Hadamard product of two vectors x and y. The core concept of an LSTM network is the cell state c_T, which behaves as the memory of the network and either allows greater preservation of past information, reducing the issues of short-term memory, or suppresses the influence of the past depending on the actions of the various gates during the training process.

LSTM networks have proven to be an effective tool in the development of reduced-order models for physical systems and have shown that they can outperform many alternate classical methods as well as other flavors of RNNs that often suffer from issues with vanishing gradients and the transmission of long-term information (Dutta et al. 2022; Ao et al. 2021). Moreover, many of the alternative popular ML and DNN-based approaches are not readily available as packages or modules inside well-known ML libraries such as TensorFlow and PyTorch. However, LSTM implementations are included as part of the core, highly efficient, GPU-accelerated modules of all these libraries. Hence, owing to the superior accessibility of training libraries and the considerable success stories of LSTM-based prediction models for dynamical systems, we have adopted it as our method of choice for modeling GWL variations over time.

In this study, the ML is constructed using two sequentially stacked LSTM layers, each comprised of 200 LSTM cells, and followed by a fully connected final layer whose output size is set to be the same as the number of time-steps needed to generate the desired model predictions. For instance, a model prediction window of 24 hr[1] with a 4 hr time-step is obtained using a fully connected final layer with (24 / 4) = 6 units.

DATA: The LSTM-based model was trained and tested on a real-world dataset consisting of various hydrometeorological state variables associated with seven different USGS gage locations in Wyoming, USA (Figure 1 and Table 1). We used USGS hourly mean groundwater level data for the period 01/01/2019 00:00 hr to 12/31/2019 23:00 hr. For our meteorological parameters, we obtained hourly ERA5 reanalysis data, which included surface temperature at

1. For a full list of the spelled-out forms of the units of measure and unit conversions used in this document, please refer to *US Government Publishing Office Style Manual*, 31st ed. (Washington, DC: US Government Publishing Office 2016), 248–52 and 345–347, respectively.*https://www.govinfo.gov/content/pkg/GPO-STYLEMANUAL-2016/pdf/GPO-STYLEMANUAL-2016.pdf*.

2 m height (2t), the clear sky direct solar radiation (ssr) at the surface, and the total precipitation (tp) (Hersbach et al. 2018).

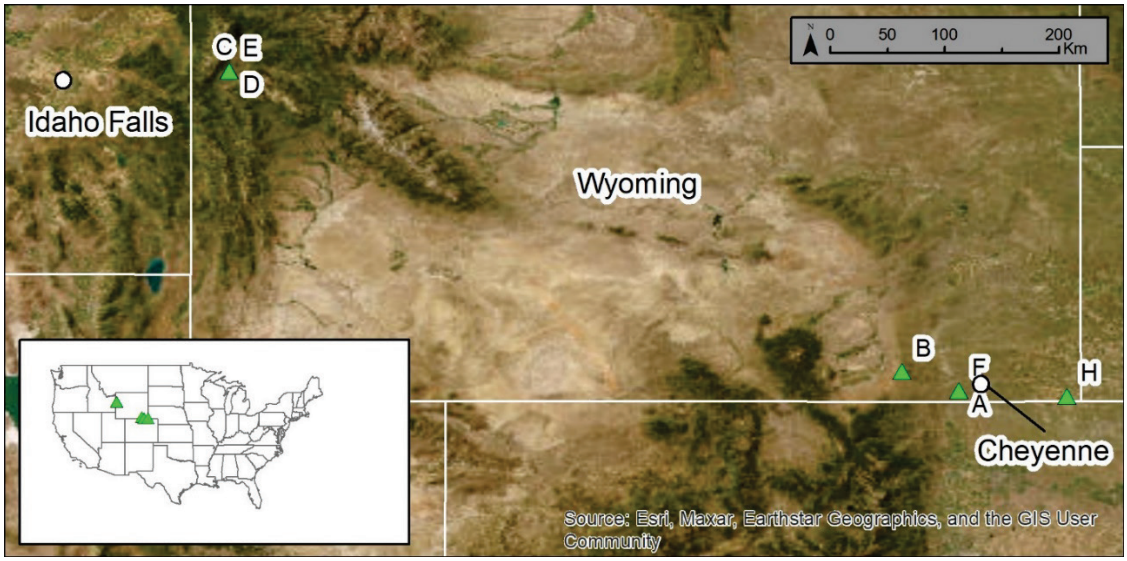

Figure 1. Study site. (Image reproduced with permission from Esri 2023. Powered by Esri.)

Table 1. USGS site information.

Station (USGS site)	Latitude	Longitude	Elevation (m)	Hydrologic Unit	Location	National Aquifer	Local Aquifer
A (410507105003802)	41.085	−105.011	2007	10190008	Laramie County	High Plains aquifer	Ogallala Formation
B (411412105272001)	41.237	−105.455	2583	10180010	Albany County	Paleozoic aquifers	Casper Formation
C (43361311044350)	43.603	−110.744	1956	17040101	Teton County	Alluvial aquifers	Terrace Deposits
D (433603110443501)	43.603	−110.744	1954	17040101	Teton County	Alluvial aquifers	Terrace Deposits
E (433551110443501)	43.597	−110.744	1953	17040101	Teton County	Alluvial aquifers	Terrace Deposits
F (410508105003801)	41.085	−105.011	2007	10190008	Laramie County	High Plains aquifer	Ogallala Formation
H (410233104093202)	41.043	−104.159	1594	10190015	Laramie County	Upper Cretaceous aquifers	Lance Formation

The dataset was split up into a training set consisting of five randomly chosen USGS gage locations (Sites A, B, C, D, and H), a validation set containing one USGS gage location (Site E), and finally, an unseen test set composed of the data corresponding to the remaining USGS gage location (Site F). For model training and evaluation purposes, all input and output variables are normalized to lie between 0 and 1. The normalized input variables at each time point are stacked together to generate a vector of dimension 4. Depending on the amount of time history used for training a particular model, a 2D array comprising of a sequence of such input vectors is fed to the model as input features both during training and evaluation. The output of the model is designed to be a sequence of scalar values representing the predicted GWLs for a desired sequence of time points in the future. The software is written in Python using the open-source TensorFlow library, and the source code is available at *https://public.git.erdc.dren.mil/RDCHLSD9/gwl_ml*. All model training and evaluation were carried out on an HP Z4 G4 workstation with an Intel Xeon(R) W-2275 CPU, an NVIDIA Quadro P1000 GPU and 64 GB RAM.

RESULTS and DISCUSSION: Figure 2 shows the pair-wise linear correlation plot between the system variables for the training set (*left plot*), the validation set (*middle plot*), and the test set (*right plot*). The surface radiation (ssr) and the air temperature at 2 m height (2t) show a positive correlation in all of the plots, as expected. The other variables do not exhibit any significant pairwise linear correlation in the training set; however, the 2 m air temperature and the GWL exhibit a strong negative correlation in both the validation set and the test set. This indicates that there is a nontrivial difference between the data distributions of the training versus the validation and test sets.

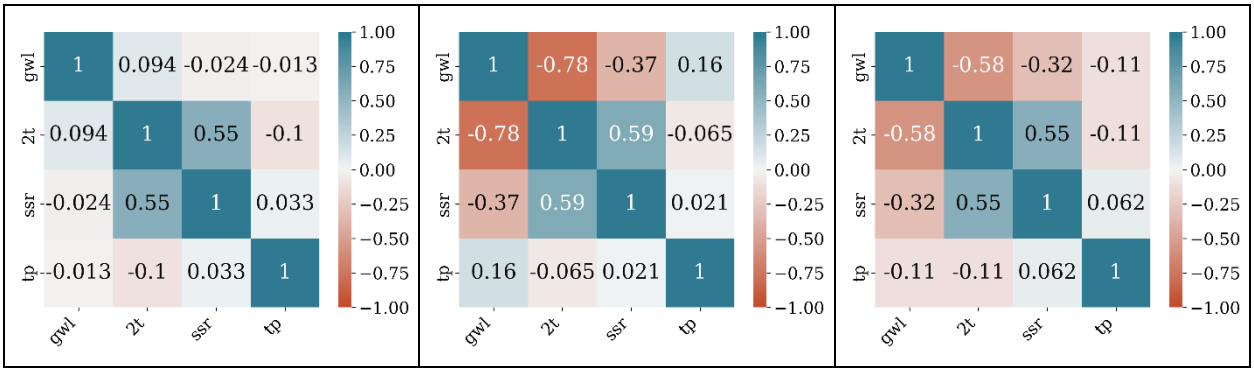

Figure 2. Correlation plots between groundwater level (GWL) and meteorological variables—2 m temperature (2t), clear sky direct solar radiation at surface (ssr), and total precipitation (tp). The subplots depict correlation values for the training dataset (*left*), validation dataset (*center*), and test dataset (*right*).

The first set of model evaluations were aimed at studying the effect of the temporal resolution of the input data on the length of the model prediction window. For this purpose, GWL predictions over gradually increasing time periods were generated from a set of models, trained with input data at an interval of 4 hr and 12 hr, and the results are presented in Figure 3 and Figure 4, respectively. The final model evaluation study, reported in Figure 5, was designed to compare the effect of the temporal resolution of the input data on the quality of GWL prediction for a fixed time period. In all of these figures, the *first two rows* compare the model predictions with the observed data for four of the training sites, and the *bottom row* shows similar comparisons for the validation Site E (*left*) and the unseen-test Site F (*right*).

Figure 3 shows the relative performance of a set of models that were trained on input data covering a 5-day period at an interval of 4 hr (i.e., 30 time-steps) and were set up to generate GWL predictions ranging from 4 hr to 120 hr (approximately 5 days) in the future, using the same time step of 4 hr. The GWL predictions for the training sites (top and middle rows) show that when the variations in the GWL levels are larger (as seen for Sites C and B), the predictive accuracy of all the models is almost identical. However, for finer-scale variations (as seen for Sites H and A), the quality of predictions deteriorates with an increase in the length of the prediction window, thus indicating that the models do not overfit the training data. The test Site F (*bottom right*), that was randomly chosen, has the smallest range of variations in the GWL levels, and thus the predictions of the trained models have more noticeable differences. Nevertheless, the results indicate that all the trained models are able to emulate the overall trend in the yearlong data; however, the fine-scale accuracy starts to gradually diminish for a prediction window longer than 24 hr.

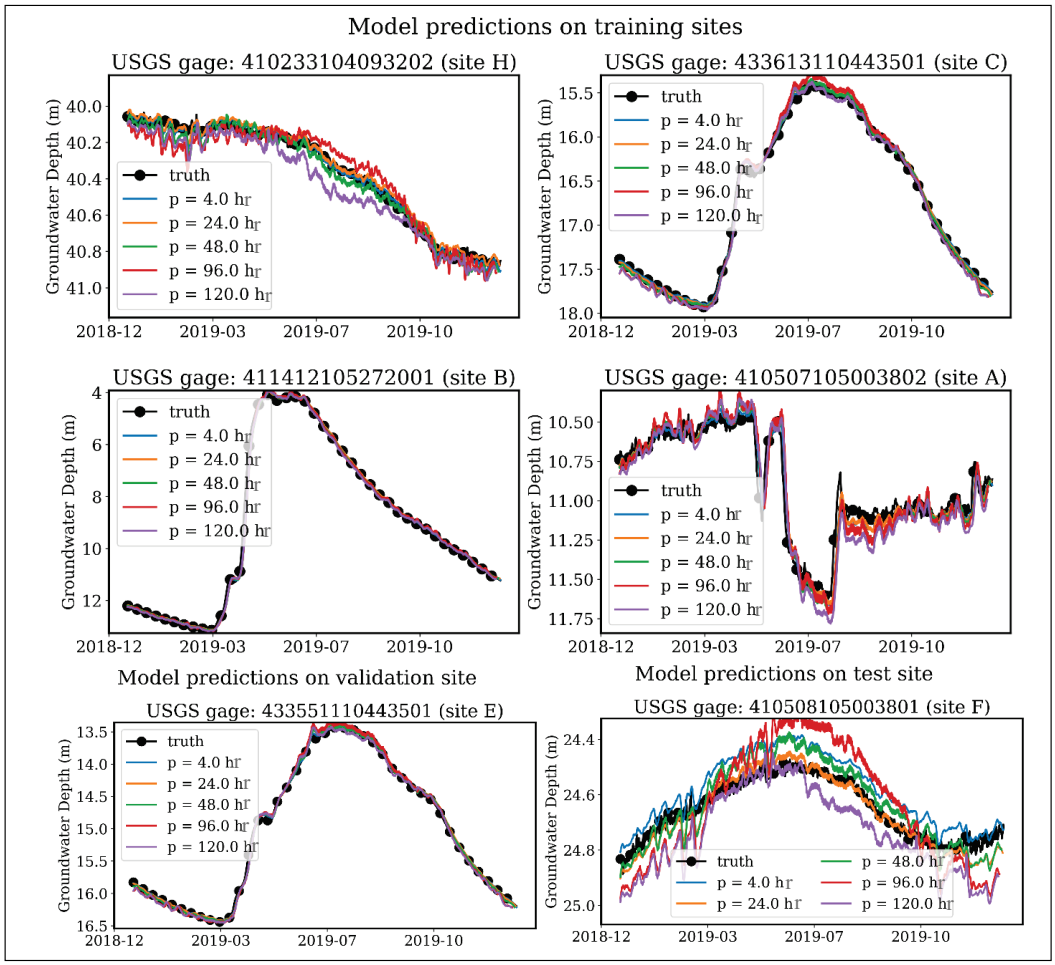

Figure 3. Comparing results obtained from long short-term memory networks (LSTM) models trained on data for 5 days at a 4 hr interval for predictions ranging from 4 hr to 120 hr. Site E is used for validation, and Site F is used for unseen testing while all the remaining sites are used for training.

Figure 4 shows the results of a study where the LSTM model was trained on 5 days of input data at an interval of 12 hr (i.e., 10 steps), and GWL predictions were generated using the same time-step for a forecast window ranging from 12 hr to 192 hr. Similar to Figure 3, it can also be observed in Figure 4 that the overall quality of model predictions, especially for sites with lower

range of variations in the GWL levels (such as Sites H, A, and F), deteriorates significantly as the temporal resolution of the input data increases. This can be expected as these sites are potentially more sensitive to daily or subdaily variations in GWL levels, but the models that are trained with lower-resolution input data (i.e., larger time-steps like 8 or 12 hr) are provided with much less information about such finer-scale variations. Hence, the overall quality of predictions of the models trained with input data at 12 hr time intervals, especially for the unseen test Site F (Figure 5), are quite significantly worse and exhibit far more spurious oscillations than those of the models trained with input data at 4 hr time intervals (Figure 3).

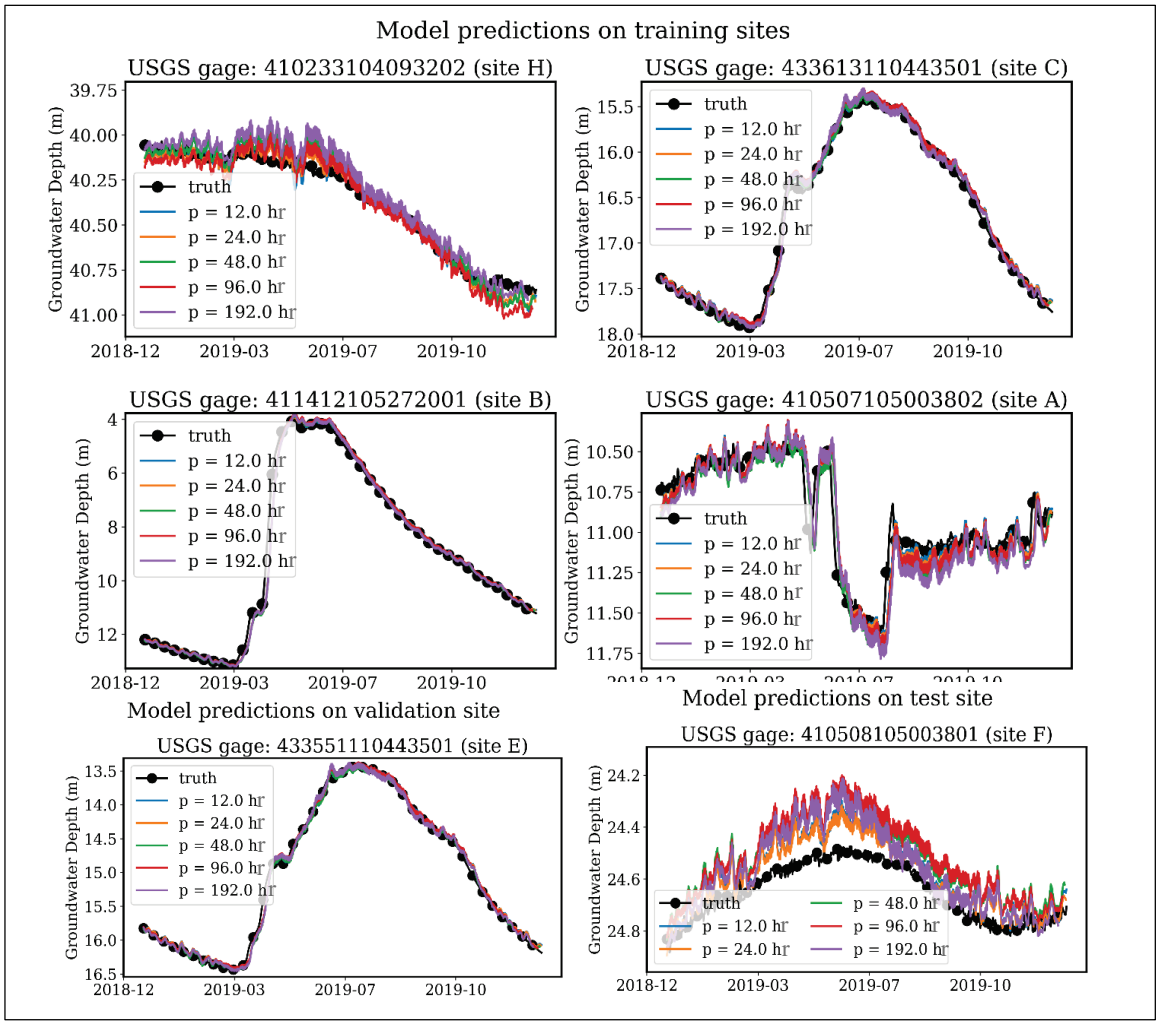

Figure 4. Comparing results obtained from LSTM models trained on data for 5 days at a 12 hr interval for predictions ranging from 12 hr to 192 hr. Site E is used for validation, and Site F is used for unseen testing while all the remaining sites are used for training.

Table 2 presents different prediction error statistics of LSTM models trained using input data with 4 hr and 12 hr time-steps. Specifically, the root mean square error (RMSE), the modified Kling-Gupta Efficiency (KGE, Kling et al. 2012), and the Spearman rank correlation coefficient (Spearman 1904) values are computed. The aggregated error values for the training sites (color coded *purple* in Table 2) are computed by taking an arithmetic mean of the corresponding error

values for all the training sites while the prediction error measures for the test set (color coded *orange*) are directly computed from the model predictions on the test site.

Table 2. Prediction error statistics of LSTM models trained with different input and output time steps (RMSE: root-mean-square error; KGE: modified Kling Gupta Efficiency; Spearman R: Spearman's rank correlation coefficient).

Training time-step = 4 hr				Training time-step = 12 hr			
Prediction length	**RMSE**	**KGE**	**Spearman R**	**Prediction length**	**RMSE**	**KGE**	**Spearman R**
p = 4 hr	0.0679	0.9773	0.9854	p = 12 hr	0.1231	0.9584	0.9552
	0.0778	0.8344	0.9752		0.0892	0.7045	0.8814
p = 24 hr	0.0800	0.9704	0.9829	p = 24 hr	0.1223	0.9632	0.9625
	0.0367	0.7895	0.9737		0.0764	0.7482	0.9187
p = 48 hr	0.0872	0.9571	0.9776	p = 48 hr	0.1281	0.9456	0.9538
	0.0581	0.6304	0.9539		0.1427	0.6404	0.8722
p = 96 hr	0.1045	0.9353	0.9570	p = 96 hr	0.1310	0.9392	0.9505
	0.1171	0.0703	0.8606		0.1509	0.5525	0.8596
p = 120 hr	0.1035	0.9446	0.9802	p = 120 hr	0.1391	0.9343	0.9489
	0.1003	0.6249	0.9733		0.116	0.5075	0.8624

The error statistics provide further quantitative evidence to support the conclusions drawn from the results in Figure 3 and Figure 4. For instance, the RMSE, KGE, and Spearman's R values are higher for the models trained on 4 hr input data than the comparable models trained on 12 hr input data. There is also a general trend of deterioration in the RMSE, KGE, and Spearman's R values as the length of the prediction window increases. However, the rate of reduction gradually diminishes as the length of the prediction window increases beyond 48 hr. For instance, using an LSTM model trained on 30 input time-steps at 4 hr intervals (or 10 input time-steps at 12 hr intervals) to predict the next 12 to 30 future time-steps (or 4 to 10 future time-steps at 12 hr intervals) can be considered a relatively data-scarce regime in which LSTM models are susceptible to overfitting and loss of generalization due to the large number of trainable model parameters. This can explain why there is a significant drop in the error statistics for the test set whereas the error statistics for the training set do not deteriorate with an increase in the length of the prediction window. Hence, for LSTM models trained on 120 hr or 5 days of past observations, the reliability of predictions can be expected to diminish for an output window longer than 48 hr. Some of these issues can be addressed with the use of regularization techniques such as dropout, weight decay, and batch normalization, or the adoption of attention mechanisms, which focus on relevant parts of the input and output sequences to enhance the information flow. However, more rigorous model training and testing needs to be done with larger data sets and longer time horizons to draw any general model design conclusions.

Figure 5 compares the predictions obtained from LSTM models trained on input data at different temporal resolutions. Each LSTM model was trained on 5 days of input data, but at four different time-steps of 2, 4, 8, and 12 hr (i.e., 60, 30, 15, and 10 steps, respectively), and GWL predictions were generated using the same respective time-steps for a 24 hr long forecast window. As observed in the previous experiments, the differences in model predictions are observed only for Sites H, A, and F, which have relatively lower ranges of GWL variations.

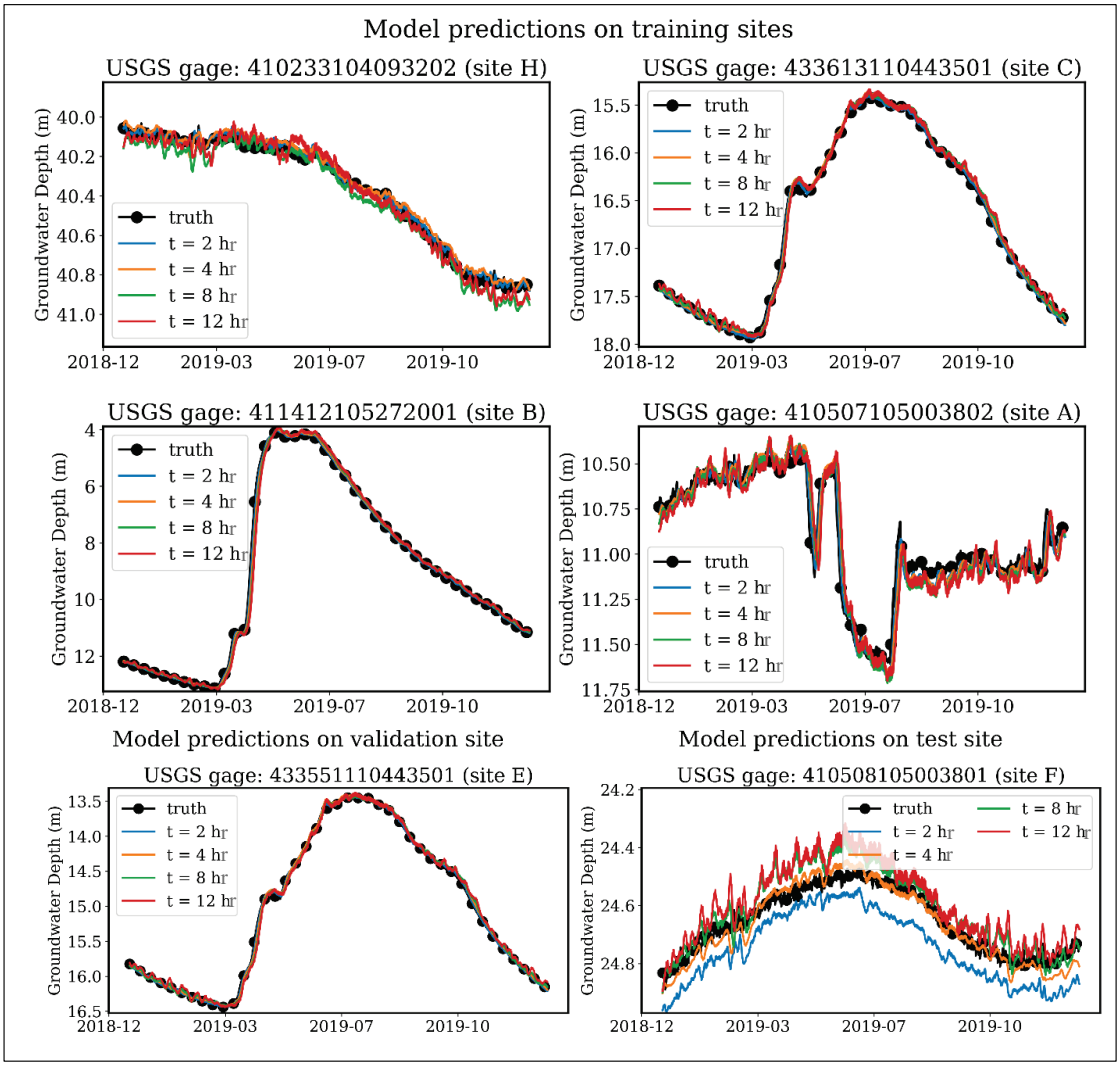

Figure 5. Comparing results obtained from LSTM models trained on data for 5 days at variable time intervals of t = 2, 4, 8, and 12 hr for future predictions of 120 hr. Site E is used for validation, and Site F is used for unseen testing while all the remaining sites are used for training.

It was observed that training an LSTM model with a 2 hr time-step posed some unique challenges. Due to the relatively longer sequence of input data (120 hr at 2 hr intervals = 60 steps), the training suffered from the well-known vanishing and exploding gradient issues, hampering the learning of long-term dependencies, and causing instability and divergence in the gradient-based optimization process. This was partially addressed by increasing the capacity of the LSTM layer from 200 cells to 400 cells, thus increasing the memory potential of the network and allowing it to extract more

relevant long-term correlations from the longer input sequence. However, this also led to more noticeable overfitting as can be observed from the disparity in the RMSE values of the 2 hr model between the training set and the test set (see Site F in Figure 5 and the top part of the second column in Table 3). The 4 hr LSTM model was found to be the most accurate and stable without a significant loss of generalization capability. The prediction accuracy gradually diminished with an increase in the time-step (i.e., decrease in temporal resolution) of the input data. In Table 3, as in Table 2, the aggregated error values for the training sites are color coded *purple*, whereas those for the test sites are color coded *orange*.

Table 3. Prediction error statistics for LSTM models trained with different input time-steps and a fixed output window length.

—	*t* = 2 hr	*t* = 4 hr	*t* = 8 hr	*t* = 12 hr
RMSE	0.0461	0.0800	0.1093	0.1223
	0.0957	0.0367	0.0606	0.0764
KGE	0.9814	0.9704	0.9600	0.9632
	0.8907	0.7895	0.7559	0.7482
Spearman R	0.9893	0.9829	0.9726	0.9625
	0.9800	0.9737	0.9461	0.9187

CONCLUSIONS AND FUTURE WORK: This ERDC technical note presents preliminary results to forecast variations in future GWL from historical observations and relevant meteorological data using an ML method known as LSTM networks. The method was trained and tested on real-world data associated with seven different USGS gage locations in Wyoming, USA, and the required groundwater and meteorological observations were obtained from USGS and the ERA5 reanalysis datasets, respectively. Our linear correlation analysis indicates that past GWL is the dominant feature in predicting the future GWL within our study region. The other variables such as temperature at a height of 2 m, total precipitation, and surface solar radiation seem to have only a weak correlation with GWL variations for our dataset. This is likely due to the semiarid climate of the chosen study region characterized by greater temperature extremes. However, an exhaustive feature-importance study needs to be conducted on an expanded set of hydrometeorological variables to conclusively determine the dominant factors affecting GWL variations. Our model predictions indicate that LSTM networks present a viable and highly versatile framework for modeling a time series of GWL variations from relevant past observation data. Moreover, the accuracy and stability of model predictions are highly sensitive to the fidelity of input data and the type as well as length of GWL forecast requirements. Therefore, any future research using this methodology should involve careful consideration of the frequency and reliability of available input data in designing the LSTM network models for desired GWL predictions.

This preliminary investigation paves the way for many promising avenues of future research. The set of meteorological input variables needs to be expanded and analyzed for their relative impact on GWL variations. The study region needs to be increased by incorporating additional USGS locations, especially across regions with diverse climatic conditions. An increase in the spatial extent of the study region will also necessitate the incorporation of topographic variables to

precisely capture the spatial patterns in GWL variations. Moreover, the capabilities of other ML algorithms need to be investigated in the time-series modeling of various hydrologic quantities of interest such as not only GWL but also soil moisture, runoff, to name a few. This will likely lead to the development of sophisticated techniques that can capture long-term trends in hydrological time series data from potentially scarce observation data.

ADDITIONAL INFORMATION: This ERDC technical note was prepared by Dr. Sourav Dutta (University of Texas at Austin), Dr. Anna Wagner (ERDC, Cold Regions Research and Engineering Laboratory, Biogeochemical Sciences Branch, Fort Wainwright, Alaska), Ms. Theadora Hall (ERDC, Information Technology Laboratory, Vicksburg, Mississippi), and Dr. Nawa Raj Pradhan (ERDC, Coastal and Hydraulics Laboratory, Hydraulic Systems Branch, Vicksburg, Mississippi) as part of ongoing research within the Intelligent Environmental Battlespace Awareness Hydrology mapping work unit's "Novel Approaches to Spatially Distributed Soil Moisture/Infiltration/Run-Off Modeling" project. In addition, appreciation is expressed to Dr. Matthew Farthing and Dr. Orie Cecil for thoughtful review of an early draft of this technical note. Questions pertaining to this technical note may be directed to Dr. Sourav Dutta (*sourav.dutta@austin.utexas.edu*).

This ERDC technical note is funded by the US Army Military Engineering Program through work item number 12JCJ0 and should be cited as follows:

> Dutta, Sourav, Anna M. Wagner, Theadora K. Hall, and Nawa Raj Pradhan. 2024. *Data-Driven Modeling of Groundwater Level Using Machine Learning*. ERDC TN-24-3. Vicksburg, MS: US Army Engineer Research and Development Center, Coastal and Hydraulics Laboratory. *http://dx.doi.org/10.21079/11681/48452*.

REFERENCES

Ao, Chang, Wenzhi Zeng, Lifeng Wu, Long Qian, Amit Kumar Srivastava, and Thomas Gaiser. 2021. "Time-Delayed Machine Learning Models for Estimating Groundwater Depth in the Hetao Irrigation District, China." *Agricultural Water Management,* 255. *https://doi.org/10.1016/j.agwat.2021.107032*.

Bear, J., M. S. Beljin, and R. R. Ross. 1992. *Ground Water Issue—Fundamentals of Ground Water Modeling*. EPA/540/S-92/005. Washington, DC: US Environmental Protection Agency.

Cai, Hejiang, Haiyun Shi, Suning Liu, and Vladan Babovic. 2021. "Impacts of Regional Characteristics on Improving the Accuracy of Groundwater Level Prediction Using Machine Learning: The Case of Central Eastern Continental United States." *Journal of Hydrology: Regional Studies* 37. *https://doi.org/10.1016/j.ejrh.2021.100930*.

Chen, Chong, Wei He, Han Zhou, Yaru Xue, and Mingda Zhu. 2020. "A Comparative Study among Machine Learning and Numerical Models for Simulating Groundwater Dynamics in the Heihe River Basin, Northwestern China." *Scientific Reports* 10 (1). *https://doi.org/10.1038/s41598-020-60698-9*.

Dutta, Sourav, Peter Rivera-Casillas, Orie M. Cecil, Matthew W. Farthing, Emma Perracchione, and Mario Putti. 2021. "Data-Driven Reduced Order Modeling of Environmental Hydrodynamics Using Deep Autoencoders and Neural Odes." *ArXiv:* 2107.02784. *http://arxiv.org/abs/2107.02784*.

Dutta, Sourav, Peter Rivera-Casillas, and Matthew William Farthing. 2021. "Neural Ordinary Differential Equations for Data-Driven Reduced Order Modeling of Environmental Hydrodynamics." *ArXiv:* 2104.13962. https://arxiv.org/abs/2104.13962.

Dutta, Sourav, Peter Rivera-Casillas, Brent Styles, and Matthew W. Farthing. 2022. "Reduced Order Modeling Using Advection-Aware Autoencoders." *Mathematical and Computational Applications* 2 7 (3): 34. https://doi.org/10.3390/mca27030034.

Esri. 2023. "World Imagery." https://www.arcgis.com/home/item.html?id=10df2279f9684e4a9f6a7f08febac2a9.

Greff, Klaus, Rupesh K. Srivastava, Jan Koutnik, Bas R. Steunebrink, and Jurgen Schmidhuber. 2017. "LSTM: A Search Space Odyssey." *IEEE Transactions on Neural Networks and Learning Systems* 28 (10): 2222–32. https://doi.org/10.1109/tnnls.2016.2582924.

Hersbach, Hans, Bill Bell, Paul Berrisford, Gionata Biavati, András Horányi, Joaquín Muñoz Sabater, Julien Nicolas, et al. 2018. "Era5 Hourly Data on Single Levels from 1979 to Present." Copernicus Climate Change Service (C3s) Climate Data Store (Cds) 10 (10.24381).https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-single-levels?tab=overview.

Hesthaven, J. S., and S. Ubbiali. 2018. "Non-Intrusive Reduced Order Modeling of Nonlinear Problems Using Neural Networks." *Journal of Computational Physics* 363: 55–78. https://doi.org/10.1016/j.jcp.2018.02.037.

Kling, Harald, Martin Fuchs, and Maria Paulin. 2012. "Runoff Conditions in the Upper Danube Basin under an Ensemble of Climate Change Scenarios." *Journal of Hydrology* 424–425: 264–77.

Mahammad, Sadik, Aznarul Islam, Pravat Kumar Shit, Abu Reza Md Towfiqul Islam, and Edris Alam. 2023. "Groundwater Level Dynamics in a Subtropical Fan Delta Region and Its Future Prediction Using Machine Learning Tools: Sustainable Groundwater Restoration." *Journal of Hydrology: Regional Studies* 47. https://doi.org/10.1016/j.ejrh.2023.101385.

Smith, Clint B., Damarys Acevedo-Acevedo, Edith Martínez-Guerra, Victor F. Medina, Michael P. Duczynski, Susan R. Wolters, Noah W. Garfinkle, Lauren Melendez, and Luisa I. Feliciano. 2022. "Developing Water Resiliency Solutions at Military Installations." *Climate Risk Management* 37. https://doi.org/10.1016/j.crm.2022.100451.

Spearman, C. 1904. "The Proof and Measurement of Association between Two Things." *The American Journal of Psychology* 15 (1): 72–101. https://doi.org/10.2307/1412159.

Wan, Zhong Yi, Pantelis Vlachas, Petros Koumoutsakos, and Themistoklis Sapsis. 2018. "Data-Assisted Reduced-Order Modeling of Extreme Events in Complex Dynamical Systems." *PloS One* 13 (5): e0197704. https://doi.org/10.1371/journal.pone.0197704.

Wu, Pin, Junwu Sun, Xuting Chang, Wenjie Zhang, Rossella Arcucci, Yike Guo, and Christopher C. Pain. 2020. "Data-Driven Reduced Order Model with Temporal Convolutional Neural Network." *Computer Methods in Applied Mechanics and Engineering 360*. https://doi.org/10.1016/j.cma.2019.112766.

Wunsch, Andreas, Tanja Liesch, and Stefan Broda. 2021. "Groundwater Level Forecasting with Artificial Neural Networks: A Comparison of Long Short-Term Memory (LSTM), Convolutional Neural Networks (CNNs), and Non-Linear Autoregressive Networks with Exogenous Input (NARX)." *Hydrology and Earth System Sciences* 25 (3): 1671–87. https://doi.org/10.5194/hess-25-1671-2021.