An Attention-based Explainable Deep Learning Approach to Spatially Distributed Hydrologic Modeling of a Snow Dominated Mountainous Karst Watershed

Qianqiu Longyang[1,5], Seohye Choi[1], Hyrum Tennant[2], Devon Hill[2], Nathan Ashmead[3], Bethany T. Neilson[2], Dennis L. Newell[4], James McNamara[3], Tianfang Xu[1*]

[1] School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ.

[2] Department of Civil and Environmental Engineering, Utah Water Research Laboratory, Utah State University, Logan, Utah.

[3] Department of Geoscience, Boise State University, Boise, Idaho.

[4] Department of Geoscience, Utah State University, Logan, Utah.

[5] Now at Kansas Geological Survey, University of Kansas, Lawrence, Kansas.

* Corresponding author: Tianfang Xu (tianfang.xu@asu.edu)

Key Points:

- An explainable spatially distributed deep learning hydrologic model is built using a spatial attention mechanism.
- The model trained with streamflow at the watershed outlet simulates discharge at subwatershed scales reasonably well.
- Recharge-discharge pathways suggested by attention weights are mostly consistent with

hydrogeochemical tracer studies.

Abstract

In many regions globally, snowmelt-recharged mountainous karst aquifers serve as crucial sources for municipal and agricultural water supplies. In these watersheds, complex interplay of meteorological, topographical, and hydrogeological factors leads to intricate recharge-discharge pathways. This study introduces a spatially distributed deep learning precipitation-runoff model that combines Convolutional Long Short-Term Memory (ConvLSTM) with a spatial attention mechanism. The effectiveness of the deep learning model was evaluated using data from the Logan River watershed and subwatersheds, a characteristically karst-dominated hydrological system in northern Utah. Compared to the ConvLSTM baseline, the inclusion of a spatial attention mechanism improved performance for simulating discharge at the watershed outlet. Analysis of attention weights in the trained model unveiled distinct areas contributing the most to discharge under snowmelt and recession conditions. Furthermore, fine-tuning the model at subwatershed scales provided insights into cross-subwatershed subsurface connectivity. These findings align with results obtained from detailed hydrogeochemical tracer studies. Results highlight the potential of the proposed deep learning approach to unravel the complexities of karst aquifer systems, offering valuable insights for water resource management under future climate conditions. Furthermore, results suggest that the proposed explainable, spatially distributed, deep learning approach to hydrologic modeling holds promise for non-karstic watersheds.

1 Introduction

Globally, and in many regions of the western U.S., karst watersheds serve as crucial

sources for municipal and agricultural water supplies. Some of these watersheds are in mountainous regions with snow-dominated hydrography, and the snowpack in these systems in the western U.S. are predicted to be adversely impacted by changing precipitation patterns due to climate change (Gergel et al., 2017; Li et al., 2017). Planning for these likely changes and effective water resource management requires accurate prediction of streamflow, which, however, is challenging for these watersheds, due to highly spatially variable snow processes and surface/subsurface heterogeneity of karst hydrogeology. Complex mountain terrain and its effect on microclimate, together, result in spatially varying snow accumulation and melt (López-Moreno et al., 2013; Sexstone and Fassnacht, 2014; Miller et al., 2022). In addition, karst systems possess intricate subsurface connectivity via sinkholes, caves, and conduits that can allow groundwater to cross basin boundaries (White, 2002; Bakalowicz, 2005). A detailed investigation of the resulting recharge-discharge pathways requires extensive field surveys that are not feasible at meso- and regional scales. As a result of their subsurface connectivity, karst watersheds are challenging for general-purpose spatially distribute hydrologic modeling frameworks such as the NOAA U.S. National Water Model (Cosgrove et al., 2024). These modeling frameworks route surface runoff (typically calculated by a land surface model) within topographically delineated watershed boundaries, and are thus unable to capture lateral groundwater flow occurring across basin boundaries typical in karst terrain.

In recent years, deep learning algorithms have emerged as an alternative approach to hydrologic modeling. Long Short-Term Memory (LSTM) networks (Hochreiter and Schmidhuber, 1997) are capable of learning temporal dynamics and have been successful for tasks such as rainfall-runoff modeling and soil moisture estimation (e.g., Fang et al., 2018; Kratzert et al., 2018). Another popular type of architecture, convolutional neural networks

(CNN), commonly used for extraction of spatial information (Fukushima, 1980; LeCun et al., 1998), has also been shown to perform well in hydrologic applications where spatial patterns are of interest (Sun et al., 2019; Pan et al., 2019; Mo et al., 2019; Anderson and Radić, 2022). More recently, the Convolutional LSTM (ConvLSTM) architecture was proposed to combine LSTM and CNN to capture spatiotemporal processes similar to movement of an object in a video (Shi et al., 2015). ConvLSTM has achieved state-of-the-art performance in tasks such as precipitation nowcasting (Shi et al., 2015), rainfall-runoff modeling (Xu et al., 2022), and streamflow forecasting (Dehghani et al., 2023; Zhu et al., 2023; Oddo et al., 2024). Unlike LSTM-based lumped rainfall-runoff models, ConvLSTM takes a spatially distributed approach and receives "image"-like inputs (e.g., gridded snowmelt, temperature). For each model grid, ConvLSTM uses the inputs (analogous to inflow) and current states at a grid and its neighbors to calculate the future state (analogous to water storage) of this grid. Although simplified, the information flow from neighbors to a given grid mimics the routing process in process-based distributed hydrologic models.

However, convolution operations use a local receptive field, thus limiting the capability of ConvLSTM to perceive long-range spatial dependencies (Lin et al., 2020). Although several ConvLSTM layers can be stacked together to represent more complex spatiotemporal dynamics occurring over long distances, using ConvLSTM alone may be insufficient to capture complex recharge-discharge pathways induced by karst geology (Xu et al., 2022). More specifically, hydrology of karst watersheds is characterized by a juxtaposition of surface runoff, slow matrix flow, and fast conduit flow. In particular, karst conduits form subsurface connectivity over long ranges and sometimes across topographically delineated basin boundaries. Therefore, in karst watersheds the performance of ConvLSTM can potentially be improved by adding global

features from the entire study area as opposed to being restricted to a small neighborhood when using convolution alone.

In parallel to the pursuit of higher prediction accuracy, there has been a long-standing interest in the hydrology community in machine learning models that are physically interpretable. A key advantage of LSTM and ConvLSTM architectures lies in their resemblance to the watershed storage, inflow and outflow dynamics (LSTM, Kratzert et al., 2019) and spatiotemporal water flow (ConvLSTM). Nevertheless, whether and how these models can reveal new insights into watershed hydrologic processes remain unclear. Several ad hoc methods are available to interpret already-trained deep learning models, such as Integrated Gradient (IG, Sundararajan et al., 2017) and Shapley Additive exPlanations (SHAP, Lundberg and Lee, 2017). However, IG tends to be sensitive to the choice of baseline, while SHAP can be computationally expensive for high dimensional problems.

In recent years, spatial attention mechanisms have attracted wide interest for improving both performance and interpretability of deep learning models. Attention is a selective cognitive process where human focuses on specific parts of information as needed, rather than processing all available information at once (Corbetta and Shulman, 2002). Selective attention allows humans to efficiently identify and concentrate on high-value information from a vast array of stimuli (Niu et al., 2021). To emulate the selective focus seen in human perception, spatial attention seeks to dynamically adjust the weights assigned to image features output by preceding layers (Guo et al., 2022). Experiments on multiple benchmark datasets showed that spatial attention was able to identify where the model should focus and improve the learned representations by promoting important features and suppressing unimportant features (Woo et al., 2018). Attention mechanisms have also been employed in hydrologic modeling, though prior

studies have predominantly focused on spatially lumped (Han et al., 2023; Wang et al., 2024) or semi-distributed modeling approaches (Feng et al., 2019; Ding et al., 2020; Feng et al., 2021). Application of attention mechanisms in fully distributed hydrologic modeling remain limited and are aimed at enhancing streamflow forecasting, relying on past streamflow data as inputs (Ghobadi and Kang, 2022). However, the potential of spatial attention as an explainable tool for understanding watershed dynamics has yet to be explored.

This study aims to present an explainable deep learning-based spatially distributed hydrologic modeling approach tailored for snow-dominated karst watersheds, while also holding promises for non-karstic counterparts. The modeling approach is demonstrated using the Logan River watershed on the Utah-Idaho border. Through a multi-scale experiment, we train an integrated ConvLSTM and spatial attention (ConvLSTM-SA) model with streamflow at watershed outlet and use it to predict discharge from subwatersheds. We further assess the capability of such a model to identify physically sensible recharge-discharge pathways within the watershed, by comparing results from interpretative analysis with hydrogeochemical analyses performed in the study watershed. We show that the spatial attention is suitable for learning karst subsurface connectivity occurring over long distances, making ConvLSTM-SA well adept at learning spatiotemporal hydrologic dynamics and potentially extendable to non-karstic watersheds. In addition, the modeling approach can serve as a screening tool to identify recharge-discharge pathways.

2 Study Area and Data

This study focuses on the canyon region of the Logan River watershed situated in the Bear River mountain range and spanning across northeastern Utah and headwaters in

132 southeastern Idaho, USA (Fig. 1). Covering an expanse of 552 km^2, the study area is
133 predominantly covered by natural land (forest, rangeland) with minimal development, and
134 through most of the study area, the Logan River is free flowing with no diversions. Over the
135 study period (1980-2022), the area experienced an average basin precipitation of approximately
136 822 mm, mostly occurring as snowfall during the winter and early spring. The watershed is
137 underlain by variably karstified carbonate formations, with the Ordovician Garden City
138 Formation and Silurian Laketown and Fish Haven dolomites as primary hosts for karst aquifer
139 development (Dover, 1995; Evans et al., 1996). Stratigraphically between the Garden City
140 Formation and Fish Haven Dolomite lies the Ordovician Swan Peak Formation, hosting shales
141 and orthoquartzites, which acts as an important aquitard influencing groundwater movement.
142 Groundwater movement is also strongly influenced by the structural geology, which is
143 dominated by the southward plunging Logan Peak syncline as well as other parallel folds and
144 numerous faults (Dover, 1995; Evans and Oaks, 1996). Rainfall and snowmelt recharge occur
145 through sinkholes, seepage along losing reaches, and diffuse infiltration into ridge slopes
146 (Spangler, 2001). Karst aquifer discharge occurs through major and minor springs within the
147 watershed, as well as direct to gaining reaches of the Logan River (Neilson et al., 2018; Lachmar
148 et al., 2021). The river primarily flows from the north and east to the south and west of the
149 watershed. However, the presence of developed karst conduits and sinkholes introduces
150 complexity to subsurface water flow direction. Previous tracer studies carried out in the western
151 portion of the watershed suggests karst piracy contributing to flow paths across topographic
152 watershed boundaries (Spangler, 2001, 2011). To accommodate this complexity, our study area
153 extends beyond the topographically delineated watershed boundary (Fig. 1).

154 Streamflow records are obtained from USGS station 10109000 located at the watershed

outlet (Fig. 1) and aggregated to daily time scale. Upstream of the USGS station, water is diverted for agricultural and municipal water uses via the Highline Canal and Dewitt Springs (Fig. 1). Daily diversion rates through the Highline Canal were acquired from USGS station 10108400. Diversion rates at Dewitt Springs were obtained from Logan City at a monthly resolution before 2020 and daily resolution since 2020. The diversion rates (monthly rates were evenly distributed to daily) were added to observed streamflow at station 10109000 to derive target data for training and validation of the deep learning model.

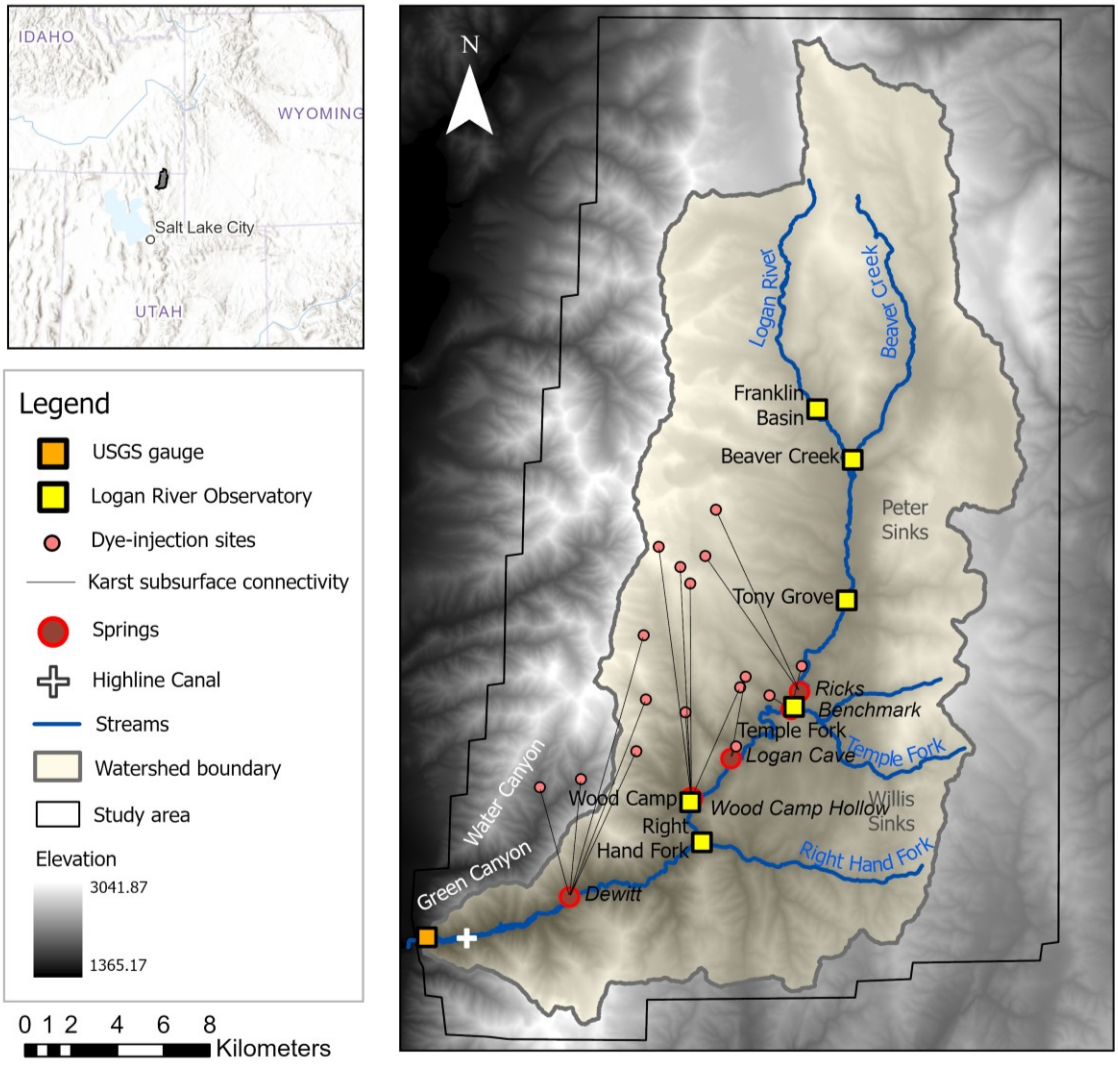

Figure 1. The Logan River watershed and its location (inset). Karst subsurface connectivity identified from previous tracer studies (Spangler, 2001, 2011) are shown by lines connecting dye-injection sites (small orange dots) and springs (large red dots). Squares show locations of streamflow gages operated by USGS (orange) and Logan River Observatory (LRO, yellow).

Snow accumulation and melt within the study area were simulated using the Utah Energy Balance (UEB) snow model based on mass and energy balances of on-land and canopy-intercepted snowpack (Mahat and Tarboton, 2012; Tarboton and Luce, 1996). We ran the UEB model using parameters described in Tyson et al. (2023) at 100 m spatial resolution and 2-hour time steps during Water Year (WY) 1981-2022. A fine spatial resolution is used to capture the spatial heterogeneity in snow accumulation and ablation processes in a mountainous terrain. The UEB model was driven by downscaled (Xu et al., 2022; Tyson et al., 2023) North American Land Data Assimilation System (NLDAS-2) Forcing dataset (Xia et al., 2012). Substantial underestimation bias was found for almost all years when comparing downscaled precipitation with observations at SNOTEL (SNOwpack TELemetry) stations, while temperature did not exhibit consistent bias patterns across years. Therefore, bias correction was applied to precipitation only (Xu et al., 2022; Tyson et al., 2023). Simulated snowmelt plus rainfall rates were then aggregated to 1.6 km-by-1.6 km resolution and daily time steps to be fed into the deep learning model. Our previous results found that this procedure was able to capture the spatial variability in snowmelt timing and rates within a 1.6 km-by-1.6 km grid, which is important for a karst watershed, at a reasonable computational expense (Xu et al., 2022).

Discharge data for Logan River Observatory (LRO) stream gages (Fig. 1) at Franklin Basin, Beaver Creek, Tony Grove, Temple Fork, Wood Camp Bridge, and Right Hand Fork were resampled from 15-minute to mean daily values (Logan River Observatory, 2024c, 2024a,

2024d, 2024f, 2024b, 2024e). The Tony Grove station has the longest period of discharge records since May 30, 2014, while other stations were installed later and have varying lengths of record. Gaps are present in the winter observations at these gages due to ice damming at the gaged cross sections, which prevents accurate reporting of streamflow.

3 Methods

3.1 Convolutional Long Short-Term Memory

The key idea of ConvLSTM is to replace the fully connected input-to-state and state-to-state transitions in classical LSTM (Hochreiter and Schmidhuber, 1997) with convolutional layers (Shi et al., 2015). This enables ConvLSTM to model dynamics that contain spatial structures. Formally, one layer of ConvLSTM can be written as:

$$
\begin{aligned}
i_t &= \sigma(W_{xi} * X_t + W_{hi} * h_{t-1} + b_i) \\
f_t &= \sigma(W_{xf} * X_t + W_{hf} * h_{t-1} + b_f) \\
g_t &= tanh\,(W_{xg} * X_t + W_{hg} * h_{t-1} + b_g) \\
o_t &= \sigma(W_{xo} * X_t + W_{ho} * X_{t-1} + b_o) \\
c_t &= f_t \odot C_{t-1} + i_t \odot g_t \\
h_t &= o_t \odot tanh\,(C_t)
\end{aligned}
\tag{1}
$$

In Eqn. (1), $X_t \in \mathbb{R}^{H \times W}$ denotes the input on time step t, $C_t, H_t \in \mathbb{R}^{C \times H \times W}$ denote cell memory and hidden state, where $H \times W$ is the spatial dimension of the study area represented by grids and C is number of channels. The hidden state and cell memory are dynamically updated

according to the input (i_t), forget (f_t), and output (o_t) gates with $W_{xi}, W_{hi}, W_{xf}, W_{hf}, W_{xg}, W_{hg}, W_{xo}, W_{ho}, b_i, b_f, b_g, b_o$ as learnable parameters. Convolutional operations are denoted as $*$, and element-wise multiplication is denoted as $\odot$. Given the relatively low amount of data from one watershed, we constrained the model complexity by using only one layer of ConvLSTM with hidden state dimension set to 20. Convolution operations are performed using 3×3 kernels and padding. In the baseline ConvLSTM-FC model, the hidden state and cell memory from the ConvLSTM layer pass through a fully connected (FC) layer, the weights and biases of which are all learnable parameters (Fig. 2a).

3.2 Spatial Attention Mechanism

We develop a ConvLSTM-SA model that combines the ConvLSTM architecture with spatial attention mechanism (Fig. 2b). Same as the ConvLSTM-FC model, the ConvLSTM-SA model has a single layer of ConvLSTM cells with 20 channels. Instead of a FC layer, ConvLSTM hidden state is processed by a spatial attention layer implemented as a modification from the Convolutional Block Attention Module (CBAM, Woo et al., 2018). CBAM was designed as an add-on module to CNNs to enhance their representation power. Let $F \in \mathbb{R}^{C \times H \times W}$ denote the feature map (i.e., output) generated by CNN, with C channels on a $H \times W$ grid, the spatial attention submodule of CBAM can be thought of as a postprocessor on F (Woo et al., 2018):

$$F' = M_s(F) \odot F, \tag{2}$$

where $F' \in \mathbb{R}^{C \times H \times W}$ is the feature map after CBAM processing, $M_s \in \mathbb{R}^{H \times W}$ is the spatial attention map, and $\odot$ denotes element-wise multiplication with M_s broadcasted along channels.

The feature map can be a concatenation of the input, cell memory, and hidden state calculated by ConvLSTM (Fig. 2), i.e., $F = [X_t; C_t; H_t]$. The spatial attention map specifies "where" (within the $H \times W$ grids) to amplify or suppress and is calculated by:

$$M_s(F) = \sigma(f^{7\times7}([AvgPool(F); MaxPool(F)]), \quad (3)$$

where σ is the sigmoid function, $f^{7\times7}$ represents a convolutional layer with a 7×7 filter for each channel, and $AvgPool$ and $MaxPool$ denote average and max, respectively, pooling operation of the feature map F across the channels.

In this study, two modifications were made to the original spatial attention mechanism as implemented in CBAM. First, we replaced the sigmoid function in Eqn. (3) with softmax function. This change ensures that the attention weights are always positive, and the sum of all weights across the entire space equals 1, providing a more physically meaningful interpretation. Essentially, this allows the original hidden state H_t to be spatially adjusted according to the attention map. Second, the spatial attention matrix was only applied to H_t, hidden state of the ConvLSTM layer:

$$H'_t = softmax(f^{7\times7}([AvgPool(F); MaxPool(F)]) \odot H_t\,. \quad (4)$$

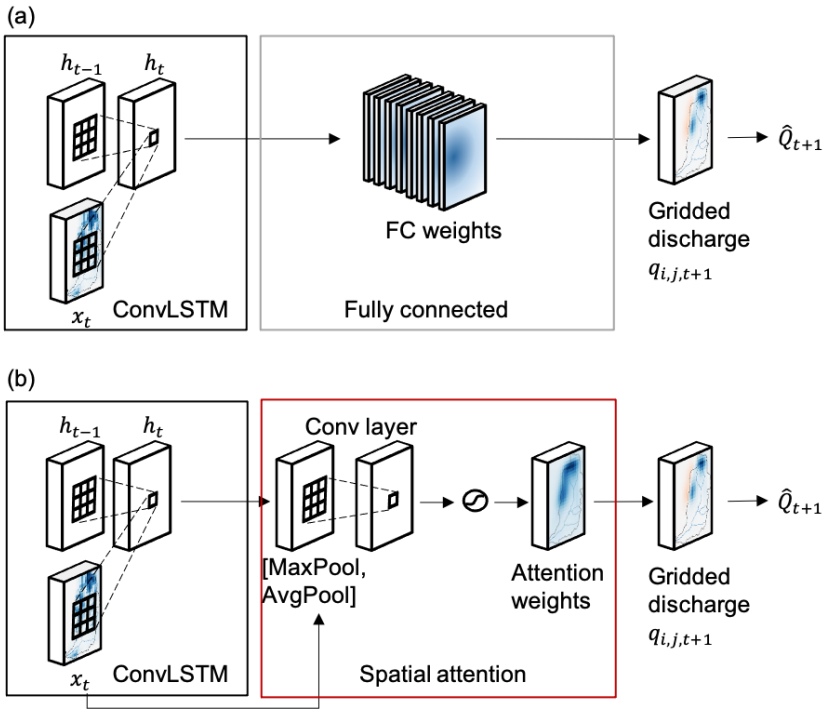

Figure 2. (a) In the baseline ConvLSTM-FC model, the hidden state from ConvLSTM layer passes through a fully connected (FC) layer before being aggregated to predict discharge at watershed outlet; (b) In the new ConvLSTM-SA model, the hidden state from the ConvLSTM layer passes through a spatial attention (SA) layer before aggregation.

3.3 A Deep Learning Approach to Spatially Distributed Hydrologic Modeling

The distributed hydrologic modeling is formulated as a sequence-to-sequence learning task. Let $X_t \in \mathbb{R}^{H \times W}$ denote the spatial distribution of snowmelt plus rainfall (simulated by UEB) over a spatial raster with a dimension of $H \times W$ at day t, and y_t denote streamflow of the following day, the task is to predict discharge (at the watershed outlet or another location along the main stem of river or its tributaries) sequence $Q = \{Q_2, Q_3 \dots, Q_{t+1}\}$, given the snowmelt plus rainfall sequence $\mathcal{X} = \{X_1, X_2, \dots, X_t\}$. The task is completed in two steps. First, the ConvLSTM-SA model calculates discharge for every grid in the model domain. More specifically, H_t, the

hidden state of the ConvLSTM layer (Eqn. 1), is processed by spatial attention layer (Eqn. 4), and then aggregated across the channels to calculate grid-wise discharge, denoted as $q_{i,j}, i = 1, \dots, H, j = 1, \dots, W$. The baseline ConvLSTM-FC model calculates grid-wise discharge similarly, except replacing SA with a FC layer. For both models, the grid-wise discharge is defined as the combined surface runoff and subsurface lateral flow (via karst conduit and/or matrix) contributed by each grid to streamflow discharge on any specific time step.

In the second step, to calculate discharge at a specified location, discharge from grids within contributing area are aggregated (summed) (Fig. 3). This process is designed to mimic a fully distributed hydrologic model. For a physically based model, gridded runoff is often routed along a river network delineated based on topography to gaging stations along the main stem river and tributaries to be comparable to observed streamflow at the gaging stations. However, in this study we chose not to perform explicit routing due to unknown subsurface connectivity. By doing so, we assume that the LSTM component in ConvLSTM-FC and ConvLSTM-SA models already capture the lag between recharge and discharge to stream, and so grid-wise discharge calculated by ConvLSTM-FC or ConvLSTM-SA represents the volume of water that would have reached the stream outlet, even though it originated from a grid at an earlier time.

For non-karst watersheds, contributing areas (watersheds and subwatersheds) are typically delineated based on topography. In karst watersheds, however, subsurface connection may result in “karst piracy”, where some groundwater flowpaths cross the surface topographic divides (White, 2002). At the watershed scale, the model domain is extended from the topographically delineated watershed boundary to account for known karst piracy identified from previous tracer studies (Spangler, 2001, 2011; Fig. 1). The ConvLSTM-FC and ConvLSTM-SA models sum up discharge from all active model grids to compute discharge at the watershed

outlet (USGS gage 10109000). This "watershed-scale" model is trained using data during WY 1981–2007 and tested for WY 2008–2022. The training configuration is detailed in Text S1, Supporting Information. The trained watershed-scale model computes $q_{i,j,t}$, grid-wise discharge at time step t, which are passed on to the subwatershed scale module.

At the subwatershed scale, subsurface connections may transfer water between adjacent subbasins. Therefore, we implemented two methods for aggregating ConvLSTM-FC or ConvLSTM-SA grid-wise discharge at the subwatershed scale. The first method creates a binary (0 and 1) mask based on topographically delineated boundary for each subwatershed; the binary masks for all subwatersheds are non-overlapping and collectively cover the entire Logan River watershed. We then use the binary masks to crop grid-wise discharge:

$$\hat{Q}_{k,t} = \sum_{i,j \in \Omega_k} q_{i,j,t}$$

where $\hat{Q}_{k,t}$ denotes computed discharge for subwatershed k at time step t, Ω_k is topographically delineated spatial extent of this subwatershed, and $q_{i,j,t}$ is ConvLSTM-FC or ConvLSTM-SA discharge for ij-th grid at time step t. We expect this method to perform well for non-karst watersheds, but will likely over- or under-estimate for a subwatershed, depending on whether it imports or exports water to neighbors. Therefore, the second method aims to inversely estimate contributing area by adding a fully-connected (FC) layer to calculate discharge for subwatershed k based on grid-wise discharge:

$$\hat{Q}_{k,t} = \sum_{i,j} w_{i,j,k} q_{i,j,t},$$

where $w_{i,j,k}, i = 1, \dots, H, j = 1, \dots, W$ are coefficients estimated using non-negative ridge regression. Let $Q_{k,t}$ denote observed discharge at time step t, we estimate $\boldsymbol{w_k} = \left[w_{1,1,k}, w_{1,2,k}, \dots, w_{i,j,k}, \dots, w_{H,W,k}\right]^T$ as

$$\underset{\boldsymbol{w_k}}{\operatorname{argmin}} \left[\sum_t \left(Q_{k,t} - \hat{Q}_{k,t}\right)^2 + \alpha \|\boldsymbol{w_k}\|_2\right],$$

$$\text{subject to } w_{i,j,k} \geq 0, \forall i, j.$$

In the above equation, α is a hyperparameter that controls the tradeoff between model goodness-of-fit to data and model complexity as represented by the L_2 regularization (also known as weight decay in machine learning context) term. The use of L_2 regularization mitigates collinearity issue (Hastie et al., 2009). In this study, collinearity exists between grid-wise discharge time series, because snowmelt plus rain time series of grids that are nearby or have similar meteorological forcing, topography, and vegetation cover are likely correlated. When discharge time series are highly correlated between two or more grids, Ridge regression tends to give similar coefficients to these time series. In contrast, Lasso, another commonly used regularized linear regression method, imposes L_1 regularization and tends to assign a high coefficient to one of the grids while zero to other grids with correlated time series; which one to receive nonzero coefficient is prone to uncertainties induced by the optimization algorithm and noise in data (Hastie et al., 2009; Zou and Hastie, 2005). Ridge regression is selected in this study to avoid false negatives, i.e., assigning a zero weight to a grid that may be contributing to discharge for a subwatershed.

The non-negative Ridge regression problem is solved using a python implementation

(Allen Institute, 2021) of the L-BFGS-B solver (Byrd et al., 1995; Zhu et al., 1997). To determine optimal hyperparameter α, we performed Ridge regression using α ranging from 0.001 to 0.5. For each value of α, regression coefficients are estimated using observed discharge at LRO gage stations (Fig. 1) during WY 2019-2022. We used data during WY 2019-2022 for training because discharge records are relatively complete during this period except gaps due to ice damming. The value of α that yielded the lowest mean square error in WY 2018 was selected as the optimal value. The performance of the final model was assessed using observed discharge during a test period of WY 2014-2017. During this test period, discharge record length varies among the LRO stations (section 2).

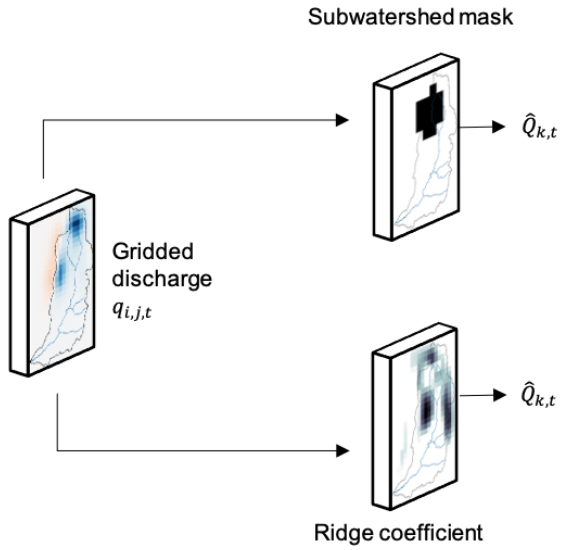

Figure 3. For a given subwatershed k, grid-wise discharge ($q_{i,j,t}$) computed by the ConvLSTM-SA model is element-wise multiplied with a binary mask based on topographic delineation (top), or coefficients determined by Ridge regression (bottom), before aggregating to calculate discharge $\hat{Q}_{k,t}$.

4 Results and Discussion

4.1 Simulating Discharge at Watershed Outlet

The performance of ConvLSTM-FC and ConvLSTM-SA models are assessed using four metrics (Table 1): percent bias (PBIAS, Gupta et al., 1999), root-mean-square error (RMSE), Nash-Sutcliff efficiency (NSE) and Kling-Gupta efficiency (KGE, Gupta et al., 2009). Compared to FC baseline, the spatial attention module improved test accuracy for all performance metrics. In addition, ConvLSTM-SA predicted hydrograph fit the observations better than FC baseline (Fig. 4). During the test period (WY2008-2022), ConvLSTM-SA achieved a KGE of 0.92, 0.96, and 0.60, respectively, during runoff (Mar. – Jun.), recession (Jul. – Oct.), and low flow (Nov. – Feb.) periods. The KGEs were all higher than KGE yielded by ConvLSTM-FC (0.86, 0.78, 0.21). Given the importance of streamflow during the recession and low flow periods for local agricultural and municipal water supply, the substantial accuracy improvement from using spatial attention is promising.

Table 1. Performance Metrics of the ConvLSTM-SA and ConvLSTM-FC Models at the Watershed Scale During Training and Test Periods. PBIAS: percent bias; RMSE: root-mean-square error; NSE: Nash-Sutcliff efficiency; KGE: Kling-Gupta efficiency.

Model	**Train / Calibrate (1981-2007)**				**Test (2008-2022)**			
	PBIAS (%)	RMSE (mm/day)	NSE	KGE	PBIAS (%)	RMSE (mm/day)	NSE	KGE
ConvLSTM-FC	0.609	0.357	0.900	0.914	2.831	0.352	0.869	0.866
ConvLSTM-SA	-3.201	0.337	0.911	0.931	0.199	0.290	0.911	0.945

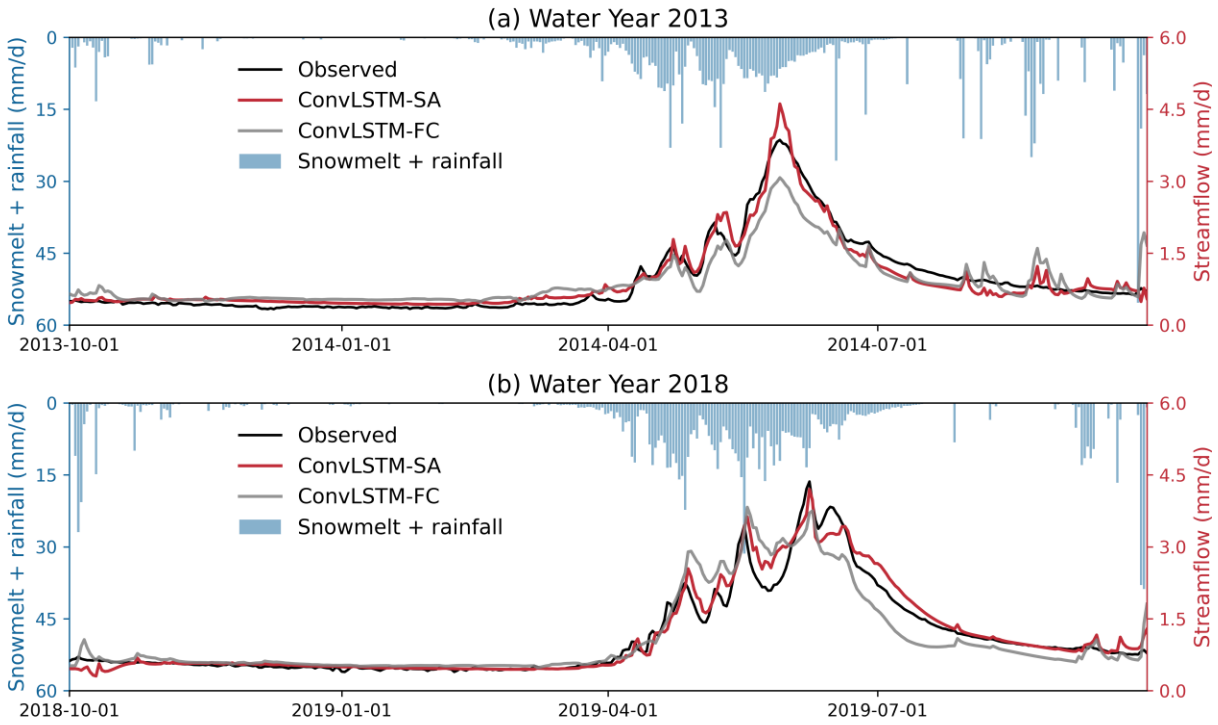

Figure 4. Spatially averaged snow plus rainfall simulated by UEB (left axes), and observed and simulated streamflow of ConvLSTM-SA and ConvLSTM-FC models (right axes) for a normal (a) and a wet (b) year. Hydrograph of the entire test period (WY2008-2022) is shown in Fig. S1, Supporting Information.

4.2 Simulating Discharge at Subwatershed Scales

At subwatershed scales, aggregating grid-wise discharge using binary masks tends to produce systematic error especially for tributaries (Table 2, Fig. 5). It also tends to yield small discharge peaks occurring in late summer (August through early October). Such peaks are learned from streamflow at the watershed outlet and are induced by summer storms. This may indicate difficulties in "deconvoluting" grid-wise discharge when recharge from rainfall does not show as much spatial variability as recharge from snowmelt. The binary mask method assumes that grids (and only these grids) within the topographic subwatershed boundary that have

contributed to discharge at the watershed outlet for a given day would contribute to discharge of this subwatershed on the same day. Thus, the difference between observed and binary mask-estimated discharge suggests the overall importing/exporting status of a subwatershed. For example, the binary mask method overestimated Beaver Creek discharge by over 300%, while underestimating Temple Fork discharge (Table 2, Fig. 5). This is consistent with findings from previous tracer studies showing karst conduit connections between large, closed basins within the Beaver Creek subwatershed to adjacent watersheds to the northwest and northeast (Figure S5, Supporting Information). Large overestimation bias is found in Logan River discharge at the Tony Grove station as the bias from Beaver Creek accumulated. The water exporting condition is also supported by tracer studies which revealed recharge-discharge pathways from high elevation areas to downstream springs (Ricks and Wood Camp Hollow, Fig. 1, Spangler, 2001; 2011); these springs contribute to a substantial portion of Logan River streamflow (Wilson, 1976). Further downstream at Wood Camp Bridge, the overestimation bias is reduced as the subwatershed area encompasses springs recharged in high elevation areas, as well as Temple Fork, that imports water from areas outside of the Logan River watershed (Figure S5, Supporting Information).

Despite the inability of the mask method to account for inter-basin karst connections, results suggest that ConvLSTM-SA with binary masks can be a promising approach to spatially distributed hydrologic modeling for non-karstic watersheds. For those watersheds, we anticipate that once trained using a downstream gage with sufficiently long streamflow records, the ConvLSTM-SA with binary masks may be able to predict streamflow at ungaged upstream locations reasonably well without the need for recalibration, especially for mesoscale watersheds. However, the accuracy of the binary mask method may deteriorate as watershed area

increases and the timing and shape of hydrograph substantially differ among subwatersheds. In such cases, it is anticipated that the deep learning model would need more training data covering a longer period to learn the "deconvoluted" grid-wise discharge using streamflow at outlet alone.

Table 2. Performance Metrics of the ConvLSTM-SA Model for Subwatersheds During WY2014-2018 (test period) Using Binary Mask and Ridge Regression Methods.

Subwatershed	Mask				Ridge Regression			
	PBIAS (%)	RMSE (mm/day)	NSE	KGE	PBIAS (%)	RMSE (mm/day)	NSE	KGE
LR Franklin Basin	69.421	0.176	0.460	0.219	-1.660	0.090	0.860	0.921
Beaver Creek	-337.288	0.422	-16.861	-3.412	32.956	0.073	0.462	0.409
LR Tony Grove	-26.923	0.351	0.346	0.377	14.619	0.178	0.831	0.709
Temple Fork	39.389	0.073	0.023	0.383	10.960	0.045	0.628	0.782
LR Wood Camp Bridge	34.925	0.868	0.180	0.498	29.660	0.814	0.279	0.558
Right Hand Fork	-9.586	0.085	0.154	0.355	25.033	0.078	0.288	0.291

On the other hand, Ridge regression substantially improved discharge simulation accuracy across four metrics for five watersheds and improving RMSE, NSE while deteriorating PBIAS and KGE for Right Hand Fork (Table 2). The metrics for Right Hand Fork may be biased due to limited availability of discharge records during the test period (<1 year). Overall, the estimated discharge appears to match well with observed hydrograph during test period for all subwatersheds. One exception was found in 2017, during which time the model underestimated a series of streamflow spikes in early spring (Fig. 5) likely induced by snowmelt events not

396 captured by the UEB model. In this year, a low bias was found in downscaled NLDAS
397 temperature, leading to subzero temperature at high elevation areas, while SNOTEL stations
398 within those areas recorded above zero temperature averaged in March and April. Therefore, the
399 UEB model substantially underestimated snowmelt rates in March and April. Given that only
400 four years of data were used for calibration and that the test period contains a larger range of
401 hydrologic conditions, the high accuracy observed here suggests ConvLSTM-SA and Ridge
402 regression to be an effective distributed hydrologic modeling approach for karst watersheds with
403 long-range subsurface connectivity.

404 ## 4.3 Interpretative Analyses

405 ### *4.3.1 Spatial attention map*

406 Tracking ConvLSTM cell memory and attention map change in time provides
407 information about watershed dynamics in different parts of water-year hydrograph. Specifically,
408 we focus on three snapshots during low flow, spring runoff, and recession periods averaged over
409 WY 1980-2022 (Fig. 6). For the study area, the lowest streamflow occurs around February 1 of
410 each year. At this time, snow is being accumulated for most of the watershed with scattered
411 snowmelt/rain at lower elevations. Around Jun. 1, snowmelt drives streamflow to peak. By
412 October 1, the snowpack has completely melted, and streamflow is sustained by groundwater,
413 with a majority from karst conduit sources (Neilson et al., 2018).

414

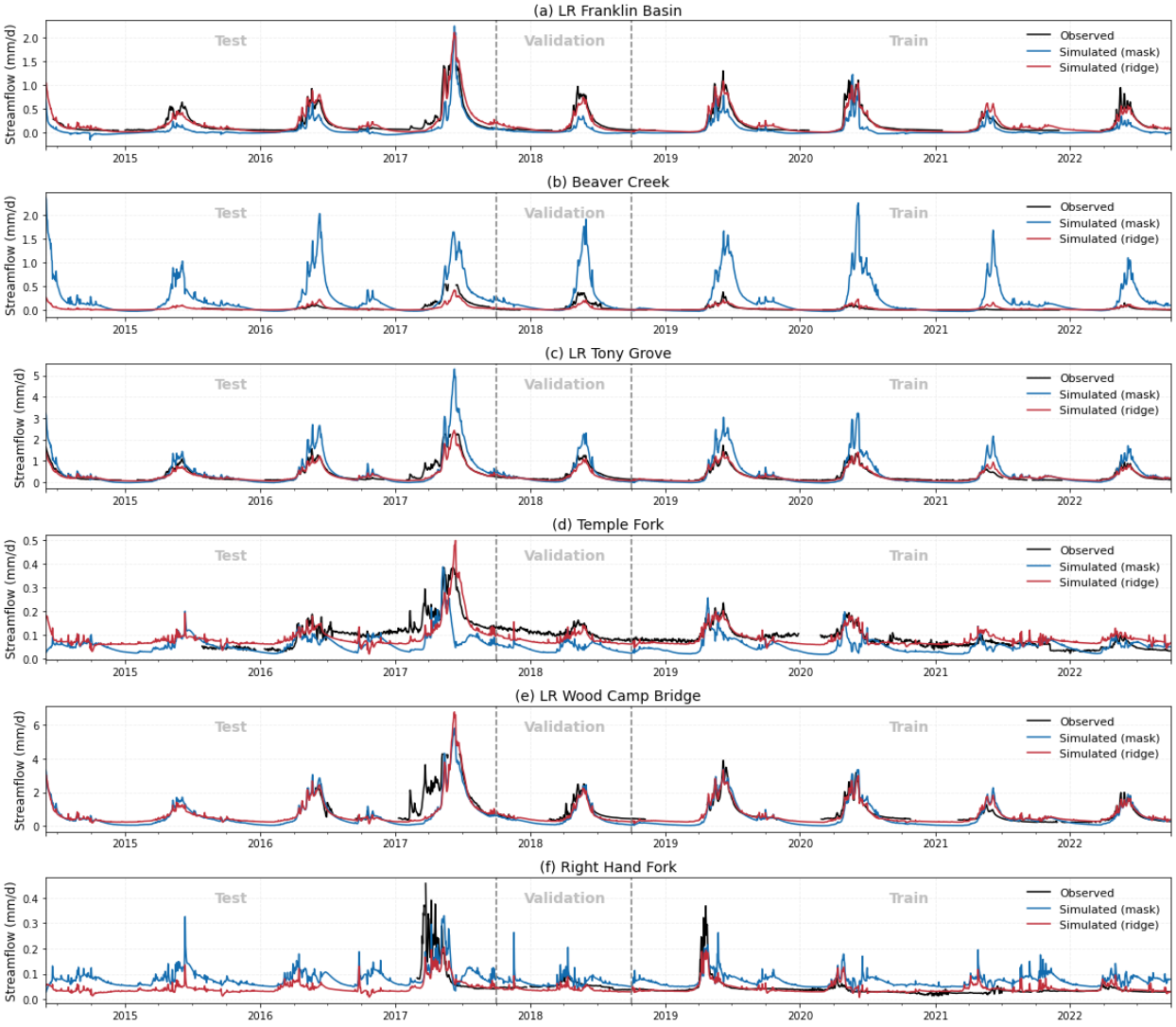

Figure 5. Observed and model estimated discharge at LRO stations along the main stem (Franklin Basin, Tony Grove, Wood Camp Bridge) and tributaries (Beaver Creek, Temple Fork, Right Hand Fork). Locations of stations are shown in Fig. 1. Discharge is aggregated using binary subwatershed masks (blue) and Ridge regression coefficients (red), respectively. Data gaps exist in observations due to differences in sensor deployment, sensor malfunction and icing events.

The cell memory of the trained ConvLSTM-SA model captures the temporal trend of water storage, which is the highest during peak flow and lower during recession and low flow periods (Fig. 6b,e,h). Meanwhile, spatial attention weights, dependent on snowmelt plus rainfall (SWIT) and cell memory, reveal discharge-generating areas and how these areas change dynamically (Fig. 6c,f,i). During low flow periods, uniform weights are observed, likely because scattered snowmelt is not sufficient to replenish depleted watershed storage and generate discharge. During spring runoff, on the other hand, high elevation snowmelt and rainfall recharge the bulk of the watershed storage, making these areas responsible for generating most of the discharge. Despite high input (snowmelt plus rain) and high cell memory, the model learned low attention weights for areas to the east of the confluence of the Logan River and Beaver Creek. Although the subwatersheds in this area are topographically part of the Logan River basin, this area has extensive karst terrain, including Peter Sinks, which has been documented as discharging towards Bear Lake to the north of the study watershed (Figure S5, Supporting Information). During the recession period, snowmelt plus rain and watershed storage exhibit different spatial patterns that together shape the spatial attention weight, which is high along the mountain ridges west of Logan River. In these areas, numerous faults and sinkholes have been found (Dover, 1995; Bahr, 2016), facilitating concentrated recharge and fast conduit flow discharging to springs along the Logan River (Fig. 1).

The consistency between the learned attention weights and local hydrogeologic information suggests the utility of the spatial attention mechanism for improving interpretability of deep learning models when sufficient data is available for training these models. Unlike ad hoc methods, including sensitivities (e.g., Anderson and Radić, 2022), that interpret already-trained deep learning models, the spatial attention module is learnable and trained

simultaneously with other components of the deep learning model. In addition, the spatial attention module can be customized to constrain the learned behavior. For example, Eqn. (4) uses the softmax function to ensure that the attention weights are positive, which also helps to constrain other learnable parameters. This led to more physically reasonable results than perturbation-based sensitivity analyses on a model without such constraints, which produced negative sensitivities of streamflow to snowmelt in our previous study (Xu et al., 2022). In this study, we inserted the spatial attention module to process hidden state. The module could also be inserted in other places within the deep neural network architecture (e.g., after the inputs), to add interpretability to desired places. However, adding the attention module at multiple places may increase data volume required to properly train the model.

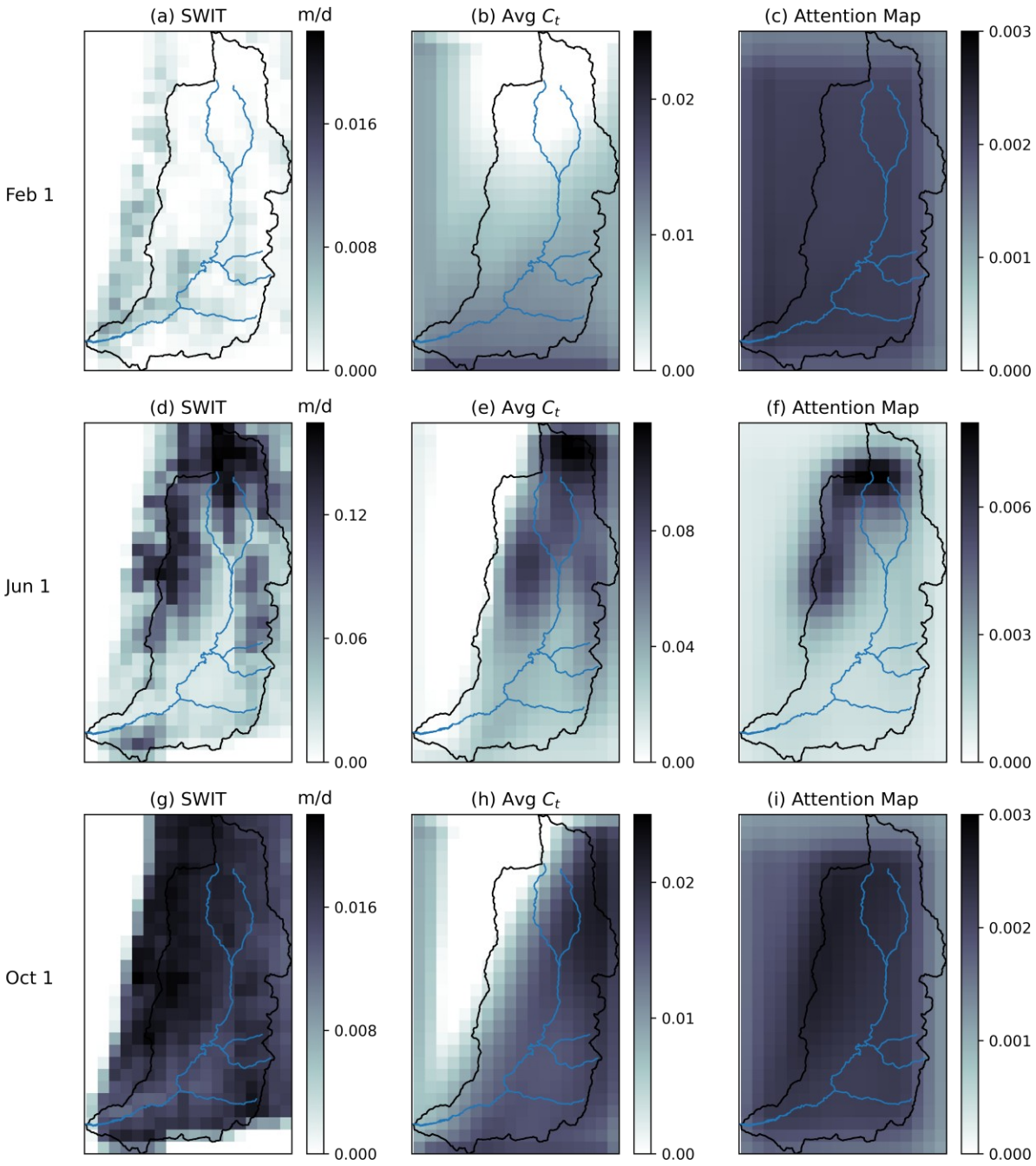

Figure 6. Multi-year average snowmelt plus rainfall (SWIT) simulated by UEB (a,d,g), ConvLSTM cell memory (C_t) averaged across channels (b,e,h), and spatial attention weights (c,f,i), on Feb. 1 (a,b,c), Jun. 1 (d,e,f), and Oct. 1 (g,h,i) of every year between WY1980-2022. Colorbar ranges differ among panels to adapt to large differences among variables shown.

4.3.2 Ridge regression coefficient map

In addition to accurately simulating subwatershed discharge, we found that Ridge regression coefficients suggest recharge-discharge pathways across subwatershed boundaries, although they may be affected by similarities between grid-wise discharge time series at different grids. To visualize such similarity, we performed principal component analysis (PCA) on a $\boldsymbol{WH} \times \boldsymbol{T}$ matrix, where $\boldsymbol{W}, \boldsymbol{H}$ are the spatial dimensions and $\boldsymbol{T}$ is number of time steps, which is the same as the length of streamflow records. Each line of the matrix corresponds to discharge time series of one grid. The leading three principal components (PCs) accounted for 74% of total variance (Fig. S2, Supporting Information). Next, a pseudo color image was generated for each gaging station (Fig. 7), such that the red, green, and blue bands of each grid are given by the contribution to discharge of this grid from the three leading PCs (Fig. S3, Supporting Information). Therefore, two grids having similar colors suggests they have similar grid-wise discharge time series, likely resulting from similar topography and climate. In the meantime, the degree of saturation (i.e., intensity) of any color in a grid in Fig. 7 is proportional to ridge regression weights of this grid, which quantifies its contribution to a given gaging station (Fig. S4, Supporting Information). Therefore, bright colors show a higher contribution than muted colors.

A grid is expected to receive a higher weight when it is contributing to discharge corresponding to a subwatershed gaging station, but not vice versa, because Ridge regression tends to give similar coefficients to grids with correlated discharge time series (section 3.3). This behavior is more suitable than regularization techniques that enforce sparsity such as Lasso. Because actual subsurface connectivity would be unknown without detailed tracer studies, we would like to identify all areas that could be contributing to subwatershed discharge to the degree

supported by data without missing potential contributing areas.

The Ridge regression weights of almost all grids are below 1 (Fig. S2, Supporting Information), which is physically reasonable. The three nested subwatersheds corresponding to main stem LRO stations show increasing weights from headwater to downstream stations. The tributary subwatersheds produced much lower streamflow than the main stem and thus receive smaller regression coefficients. Similar spatial patterns were found between Ridge coefficients of Logan River (LR) - Franklin Basin (Fig. 7a) and Beaver Creek (Fig. 7b). Hydrogeochemical data suggests a small portion of Franklin Basin discharge originates from Beaver Creek (Ashmead et al., 2023). However, a similarity is noticeable between grid-wise discharge time series of grids receiving high regression coefficients for the two subwatersheds, suggesting that at least some of the high weights may be a false positive. For Temple Fork and Right Hand Fork, the method assigns moderate coefficients to headwaters of the two subwatersheds, but also picks up areas west of the river and south of the study area, which are likely to be false positives due to collinearity. However, high coefficients assigned to the east bank from Beaver Creek to Temple Fork coincide with the Temple Ridge Fault and may suggest subsurface connectivity given the highly karstified terrain in that area (Dover, 1995).

The above results underscore the potential of our modeling approach (ConvLSTM-SA complemented by Ridge regression and PCA) to serve as a screening tool for possible contributing areas that do not follow topographic subbasin boundaries or a method for anticipating locations of karst piracy. For areas with distinct grid-wise discharge signatures, as revealed by PCA, a high Ridge regression weight is a relatively strong indicator of contributing area, while false positives are possible for areas with correlated grid-wise discharge signature. Based on the screening results, field campaign and tracer studies can be designed to collect data

to rule out false positives and establish true recharge-discharge pathways.

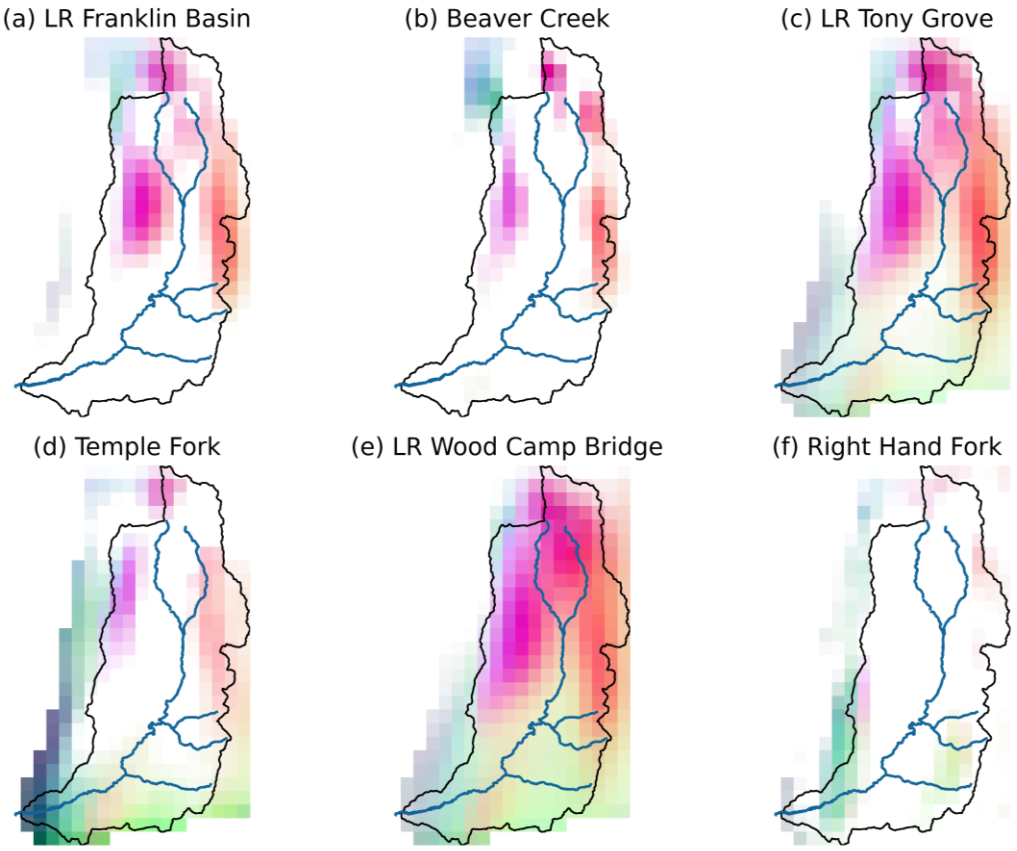

Figure 7. Pseudo color rendering of Ridge regression coefficients estimated using discharge at LRO stations along the main stem (Franklin Basin, Tony Grove, Wood Camp Bridge) and tributaries (Beaver Creek, Temple Fork, Right Hand Fork). Locations of stations are shown in Fig. 1. Regression weights are shown in Fig. S2, and pseudo color is determined by principal component analysis (Fig. S3, Supporting Information).

5 Conclusions

This study developed an explainable, spatially distributed, deep learning-based approach to hydrologic modeling in a snow-dominated mountainous karst watershed, leveraging the power of Convolutional Long Short-Term Memory (ConvLSTM) integrated with a spatial attention

mechanism. The efficacy of the approach was demonstrated through a case study focused on the Logan River watershed. Compared to the baseline ConvLSTM model, spatial attention improved simulation accuracy of discharge at the watershed outlet during the test period. In addition, the spatial attention weights computed by the trained model revealed key areas contributing to discharge under low flow, recession, and runoff periods, aligning well with known hydrogeological features and previous hydrogeochemical tracer studies.

Next, the model trained using discharge at the watershed outlet was applied to subwatershed scales. When the model predicted grid-wise discharge was aggregated by topographically delineated contributing areas, bias was observed in aggregated discharge and suggests cross-basin water transfers. Simulation accuracy of subwatershed discharges is significantly enhanced by the use of Ridge regression. Comparison between Ridge regression weights and known hydrogeologic connections shows potential of Ridge regression as a screening tool for possible recharge-discharge pathways of karst watersheds.

The presented approach proves adept at capturing the complex spatiotemporal dynamics of a mountainous karst watershed. This work not only enhances our ability to predict hydrological responses in these challenging environments, but also contributes to the broader field of hydrologic modeling, because the ConvLSTM-SA model can also be used as a spatially distributed hydrologic model for non-karst watersheds. Once trained on a downstream gage, the ConvLSTM-SA with binary masks can potentially predict streamflow at ungaged upstream locations. When upstream gages are available, observed subwatershed discharge can be utilized with Ridge regression to infer inter-basin connections. Future research should focus on extending this modeling approach to more diverse datasets of mountainous karst systems and testing the approach's applicability to non-karstic watersheds and at larger scales.

Acknowledgments

This work was supported by the US National Science Foundation Hydrologic Sciences program grants 2043150/2043363/2044051. The authors are grateful for the thoughtful review and suggestions by Sean A. McKenna, an anonymous reviewer, and the Associate Editor.

Open Research

All data used in this research are publicly available. The UEB software is available via Tarboton et al., (2015). The data and code used for simulating streamflow are available at Longyang et al. (2024).

References

Anderson, S., & Radić, V. (2022). Evaluation and interpretation of convolutional long short-term memory networks for regional hydrological modelling. *Hydrology and Earth System Sciences*, *26*(3), 795-825.

Allen Institute. (2021). Nonnegative Ridge Regression. Available from https://github.com/AllenInstitute/mouse_connectivity_models/tree/master/mcmodels/regressors/nonnegative_linear.

Bahr, K. (2016). *Structural and Lithological Influences on the Tony Grove Alpine Karst System, Bear River Range, North Central Utah*.

Bakalowicz, M. (2005). Karst groundwater: a challenge for new resources. *Hydrogeology journal*, *13*, 148-160.

Byrd, R. H., Lu, P., Nocedal, J., & Zhu, C. (1995). A limited memory algorithm for bound constrained optimization. *SIAM Journal on scientific computing*, *16*(5), 1190-1208. Access via https://epubs.siam.org/doi/epdf/10.1137/0916069.

Corbetta, M., & Shulman, G. L. (2002). Control of goal-directed and stimulus-driven attention in the brain. *Nature reviews neuroscience*, *3*(3), 201-215.

Cosgrove, B., Gochis, D., Flowers, T., Dugger, A., Ogden, F., Graziano, T., Clark, E., Cabell, R., Casiday, N., Cui, Z. and Eicher, K., (2024). NOAA's National Water Model: Advancing operational hydrology through continental-scale modeling. *JAWRA Journal of the American Water Resources Association*, *60*(2), pp.247-272.

Dehghani, A., Moazam, H.M.Z.H., Mortazavizadeh, F., Ranjbar, V., Mirzaei, M., Mortezavi, S., Ng, J.L. and Dehghani, A., (2023). Comparative evaluation of LSTM, CNN, and ConvLSTM for hourly short-term streamflow forecasting using deep learning approaches. *Ecological Informatics*, *75*, p.102119.

Ding, Y., Zhu, Y., Feng, J., Zhang, P., & Cheng, Z. (2020). Interpretable spatio-temporal attention LSTM model for flood forecasting. *Neurocomputing*, *403*, 348-359.

Dover, J.H. (1995). Geologic map of the Logan 30' x 60' quadrangle, Cache and Rich Counties,

Utah, and Lincoln and Uinta Counties, Wyoming: U.S. Geological Survey Miscellaneous Investigations Series Map I-2210, 1 pl., scale 1:100,000.

Evans, J. P., & Oaks Jr, R. Q. (1996). Three-dimensional variations in extensional fault shape and basin form: The Cache Valley basin, eastern Basin and Range province, United States. *Geological Society of America Bulletin*, *108*(12), 1580-1593.

Evans, J. P., McCalpin, J. P., & Holmes, D. C. (1996). Geologic Map of the Logan Quadrangle, Cache County, Utah.

Fang, K., Pan, M., & Shen, C. (2018). The value of SMAP for long-term soil moisture estimation with the help of deep learning. *IEEE Transactions on Geoscience and Remote Sensing*, *57*(4), 2221-2233.

Feng, J., Wang, Z., Wu, Y., & Xi, Y. (2021, July). Spatial and temporal aware graph convolutional network for flood forecasting. In *2021 International joint conference on neural networks (IJCNN)* (pp. 1-8). IEEE.

Feng, J., Yan, L., & Hang, T. (2019). Stream-flow forecasting based on dynamic spatio-temporal attention. *IEEE Access*, *7*, 134754-134762.

Fukushima, K. (1980). Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, *36*(4), 193-202.

Gergel, D. R., Nijssen, B., Abatzoglou, J. T., Lettenmaier, D. P., & Stumbaugh, M. R. (2017). Effects of climate change on snowpack and fire potential in the western USA. *Climatic Change*, *141*, 287-299.

Ghobadi, F., & Kang, D. (2022). Improving long-term streamflow prediction in a poorly gauged basin using geo-spatiotemporal mesoscale data and attention-based deep learning: A comparative study. *Journal of Hydrology*, *615*, 128608.

Guo, M.H., Xu, T.X., Liu, J.J., Liu, Z.N., Jiang, P.T., Mu, T.J., Zhang, S.H., Martin, R.R., Cheng, M.M. and Hu, S.M., (2022). Attention mechanisms in computer vision: A survey. *Computational visual media*, *8*(3), pp.331-368.

Gupta, S. K., Ritchey, N. A., Wilber, A. C., Whitlock, C. H., Gibson, G. G., & Stackhouse Jr, P. W. (1999). A climatology of surface radiation budget derived from satellite data. *Journal of climate*, *12*(8), 2691-2710.

Gupta, H. V., Kling, H., Yilmaz, K. K., & Martinez, G. F. (2009). Decomposition of the mean squared error and NSE performance criteria: Implications for improving hydrological modelling. *Journal of hydrology*, *377*(1-2), 80-91.

Han, D., Liu, P., Xie, K., Li, H., Xia, Q., Cheng, Q., Wang, Y., Yang, Z., Zhang, Y. and Xia, J., (2023). An attention-based LSTM model for long-term runoff forecasting and factor recognition. *Environmental Research Letters*, *18*(2), p.024004.

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). *The elements of statistical learning: data mining, inference, and prediction* (Vol. 2, pp. 1-758). New York: springer.

Hochreiter, S., & Schmidhuber, J. (1997). Long short-term memory. *Neural computation*, *9*(8), 1735-1780.

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), 6005-6022.

Kratzert, F., Herrnegger, M., Klotz, D., Hochreiter, S., Klambauer, G. (2019). NeuralHydrology–interpreting LSTMs in hydrology. *Explainable AI: Interpreting, explaining and visualizing deep* learning, 347–362.

Lachmar, T., Skyler, S., & Newell, D. (2021). Geochemical insights into groundwater movement

in alpine karst, Bear River Range, Utah, USA. *Hydrogeology Journal*, *29*(2), 687-701.
LeCun, Y., Bottou, L., Bengio, Y., & Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, *86*(11), 2278-2324.
Li, D., Wrzesien, M. L., Durand, M., Adam, J., & Lettenmaier, D. P. (2017). How much runoff originates as snow in the western United States, and how will that change in the future?. *Geophysical Research Letters*, *44*(12), 6163-6172.
Lin, Z., Li, M., Zheng, Z., Cheng, Y., & Yuan, C. (2020, April). Self-attention convlstm for spatiotemporal prediction. In *Proceedings of the AAAI conference on artificial intelligence* (Vol. 34, No. 07, pp. 11531-11538).
Logan River Observatory. (2024a). Logan River Observatory: Beaver Creek above confluence with Logan River Aquatic Site (BC_CONF_A) Quality Controlled Data | CUAHSI HydroShare. https://www.hydroshare.org/resource/d680a2af2ee6491285381fb84d45c871/. Accessed: June 2023.
Logan River Observatory. (2024b). Logan River Observatory: Logan River at Wood Camp Bridge Aquatic Site (LR_WCB_A) Quality Controlled Data | CUAHSI HydroShare. https://www.hydroshare.org/resource/b8f196490b6348b7b2945ef559924f87/. Accessed: June 2023.
Logan River Observatory. (2024c). Logan River Observatory: Logan River near Franklin Basin Aquatic Site (LR_FB_BA) Quality Controlled Data | CUAHSI HydroShare. https://www.hydroshare.org/resource/1bb3210918414e13b077e87798d4a696/. Accessed: June 2023.
Logan River Observatory. (2024d). Logan River Observatory: Logan River near Tony Grove Aquatic Site (LR_TG_BA) Quality Controlled Data | CUAHSI HydroShare. https://www.hydroshare.org/resource/b93121c191a94abbb288acabba07f954/. Accessed: June 2023.
Logan River Observatory. (2024e). Logan River Observatory: Right Hand Fork above confluence with Logan River Aquatic Site (RHF_CONF_A) Quality Controlled Data | CUAHSI HydroShare. https://www.hydroshare.org/resource/a017b44b64804311abaeb9d9917e8fcf/. Accessed: June 2023.
Logan River Observatory. (2024f). Logan River Observatory: Temple Fork above confluence with Logan River Aquatic Site (TF_CONF_A) Quality Controlled Data | CUAHSI HydroShare. https://www.hydroshare.org/resource/499bd5326b1443b29c9ac75b2903a025/. Accessed: June 2023.
Longyang, Q., Choi, S., Tennant, H., Hill, D., Ashmead, N., Neilson, B. T., Newell, D. L., McNamara, J., & Xu, T. (2024). ConvLSTM-SA model code for Logan River Basin (LRB) (0.0.1). Zenodo. https://doi.org/10.5281/zenodo.11094821
López-Moreno, J.I., Fassnacht, S.R., Heath, J.T., Musselman, K.N., Revuelto, J., Latron, J., Morán-Tejeda, E. & Jonas, T., (2013). Small scale spatial variability of snow density and depth over complex alpine terrain: Implications for estimating snow water equivalent. *Advances in water resources*, *55*, pp.40-52.
Lundberg, S. M., & Lee, S. I. (2017). A unified approach to interpreting model predictions. *Advances in neural information processing systems*, *30*.
Mahat, V., & Tarboton, D. G. (2012). Canopy radiation transmission for an energy balance snowmelt model. *Water Resources Research*, *48*(1).
Miller, Z. S., Peitzsch, E. H., Sproles, E. A., Birkeland, K. W., & Palomaki, R. T. (2022).

668 Assessing the seasonal evolution of snow depth spatial variability and scaling in complex
669 mountain terrain. *The Cryosphere*, *16*(12), 4907-4930.
670 Mo, S., Zabaras, N., Shi, X., & Wu, J. (2019). Deep autoregressive neural networks for high-
671 dimensional inverse problems in groundwater contaminant source identification. *Water*
672 *Resources Research*, *55*(5), 3856-3881.
673 Neilson, B.T., Tennant, H., Stout, T.L., Miller, M.P., Gabor, R.S., Jameel, Y., Millington, M.,
674 Gelderloos, A., Bowen, G.J. & Brooks, P.D., (2018). Stream centric methods for determining
675 groundwater contributions in karst mountain watersheds. *Water Resources Research*, *54*(9),
676 pp.6708-6724.
677 Niu, Z., Zhong, G., & Yu, H. (2021). A review on the attention mechanism of deep
678 learning. *Neurocomputing*, *452*, 48-62.
679 Oddo, P. C., Bolten, J. D., Kumar, S. V., & Cleary, B. (2024). Deep Convolutional LSTM for
680 improved flash flood prediction. *Frontiers in Water*, *6*, 1346104.
681 Pan, B., Hsu, K., AghaKouchak, A., & Sorooshian, S. (2019). Improving precipitation estimation
682 using convolutional neural network. *Water Resources Research*, *55*(3), 2301-2321.
683 Sundararajan, M., Taly, A., & Yan, Q. (2017, July). Axiomatic attribution for deep networks.
684 In *International conference on machine learning* (pp. 3319-3328). PMLR.
685 Sexstone, G. A., & Fassnacht, S. R. (2014). What drives basin scale spatial variability of
686 snowpack properties in northern Colorado?. *The Cryosphere*, *8*(2), 329-344.
687 Shi, X., Chen, Z., Wang, H., Yeung, D. Y., Wong, W. K., & Woo, W. C. (2015). Convolutional
688 LSTM network: A machine learning approach for precipitation nowcasting. *Advances in*
689 *neural information processing systems*, *28*.
690 Spangler, L. E. (2001). Delineation of recharge areas for karst springs in Logan Canyon, Bear
691 River Range, northern Utah. *US Geol Surv Water Resour Invest Rep*, *1*, 186-193.
692 Spangler, L. E. (2011). Karst hydrogeology of the Bear River Range in the vicinity of the Logan
693 River, Northern Utah. In *Geological Society of America Rocky Mountain-Cordilleran Section*
694 *Meeting, US Geological Survey*.
695 Sun, A. Y., Scanlon, B. R., Zhang, Z., Walling, D., Bhanja, S. N., Mukherjee, A., & Zhong, Z.
696 (2019). Combining physically based modeling and deep learning for fusing GRACE satellite
697 data: can we learn from mismatch?. *Water Resources Research*, *55*(2), 1179-1195.
698 Tarboton, D. G., & Luce, C. H. (1996). *Utah energy balance snow accumulation and melt model*
699 *(UEB)* (p. 63). Utah Water Research Laboratory.
700 Tarboton, D. G., Gichamo, T. Z., & Merck, M. (2015). UEB [Source code]. GitHub.
701 https://github.com/dtarb/UEB.
702 Tyson, C., Longyang, Q., Neilson, B. T., Zeng, R., & Xu, T. (2023). Effects of meteorological
703 forcing uncertainty on high-resolution snow modeling and streamflow prediction in a
704 mountainous karst watershed. *Journal of Hydrology*, *619*, 129304.
705 Wang, H., Qin, H., Liu, G., Huang, S., Qu, Y., Qi, X., & Zhang, Y. (2024). Hierarchical attention
706 network for short-term runoff forecasting. *Journal of Hydrology*, 131549.
707 White, W. B. (2002). Karst hydrology: recent developments and open questions. *Engineering*
708 *geology*, *65*(2-3), 85-105.
709 Wilson, J. R. (1976). Glaciated dolomite karst in the Bear River Range, Utah. *PhD dissertation,*
710 *University of Utah, Department of Geology and Geophysics*.
711 Woo, S., Park, J., Lee, J. Y., & Kweon, I. S. (2018). Cbam: Convolutional block attention
712 module. In *Proceedings of the European conference on computer vision (ECCV)* (pp. 3-19).
713 Xia, Y., Mitchell, K., Ek, M., Sheffield, J., Cosgrove, B., Wood, E., et al. (2012). Continental-

scale water and energy flux analysis and validation for the North American land data assimilation system project phase 2 (NLDAS-2): 1. Intercomparison and application of model products. Journal of Geophysical Research, 117(D3). https://doi.org/10.1029/2011jd016048

Xu, T., Longyang, Q., Tyson, C., Zeng, R., & Neilson, B. T. (2022). Hybrid physically based and deep learning modeling of a snow dominated, mountainous, karst watershed. *Water Resources Research*, *58*(3), e2021WR030993.

Zhu, C., Byrd, R. H., Lu, P., & Nocedal, J. (1997). Algorithm 778: L-BFGS-B: Fortran subroutines for large-scale bound-constrained optimization. *ACM Transactions on mathematical software (TOMS)*, *23*(4), 550-560.

Zhu, S., Wei, J., Zhang, H., Xu, Y., & Qin, H. (2023). Spatiotemporal deep learning rainfall-runoff forecasting combined with remote sensing precipitation products in large scale basins. *Journal of Hydrology*, *616*, 128727.

Zou, H., & Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society Series B: Statistical Methodology*, *67*(2), 301-320.