# The Role of Snow Information in Deep Learning for Sub-Seasonal Streamflow Forecasting

Hrishant Shiwakoti Sharma[1], Eunsang Cho[2], Hyunglok Kim[3]

[1]Department Of Computer Science, Texas State University, San Marcos, TX, USA, [2]Ingram School of Engineering, Texas State University, San Marcos, TX, USA

[3]School of Earth Sciences and Environmental Engineering, Gwangju Institute of Science and Technology (GIST), South Korea

Email: pfj19@txstate.edu

## INTRODUCTION

- Reliable streamflow forecasting is critical for water supply outlooks, flood management, reservoir operations, and ecological planning, particularly in snow-dominated mountain basins (Koster et al., 2010).
- Traditional physics-based models still struggle to provide reliable predictions in these regions. Recent advances in deep learning offer powerful tools for learning nonlinear hydrologic relationships from historical data (Ng et al., 2023).
- This study demonstrates how incorporating snow information improves sub-seasonal streamflow forecasting using a hybrid CNN–LSTM model, along with other deep learning approaches, in the Yellowstone River Basin.

## RESEARCH QUESTIONS

- Which deep learning architectures perform best for sub-seasonal streamflow forecasting in snow-dominated basins?
- How accurately can deep learning models capture both the magnitude and timing of peak streamflow at 7, 14, and 30-day lead times?
- Does the contribution of SWE on streamflow prediction increase with forecast lead times?

## DATA OVERVIEW

| Features | Temporal Resolution | Spatial Resolution | Data Source | Period |
|---|---|---|---|---|
| Flow | Daily | Point | USGS | 1981-2023 |
| Temperature | Daily | 4km | PRISM | 1981-2023 |
| Precipitation | Daily | 4km | PRISM | 1981-2023 |
| SWE | Daily | 4km | UA SWE | 1981-2023 |

- Daily observed discharge data from United States Geological Survey (USGS) were used as prediction target.
- Daily precipitation and temperature from Parameter-elevation Regressions on Independent Slopes Model (PRISM) representing climate forcing.
- Daily University of Arizona Snow Water Equivalent (UA SWE) data used to capture snowpack storage and melt effects

## WORKFLOW AND METHODOLOGY

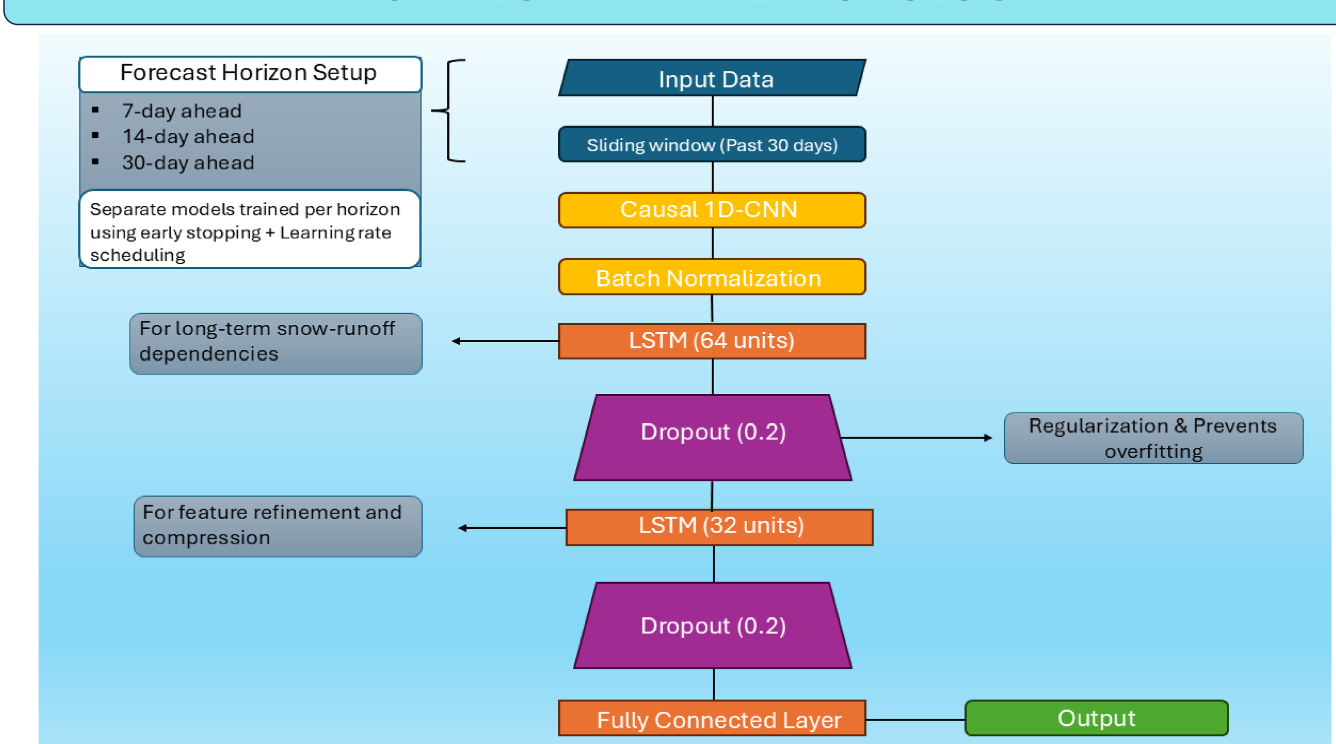
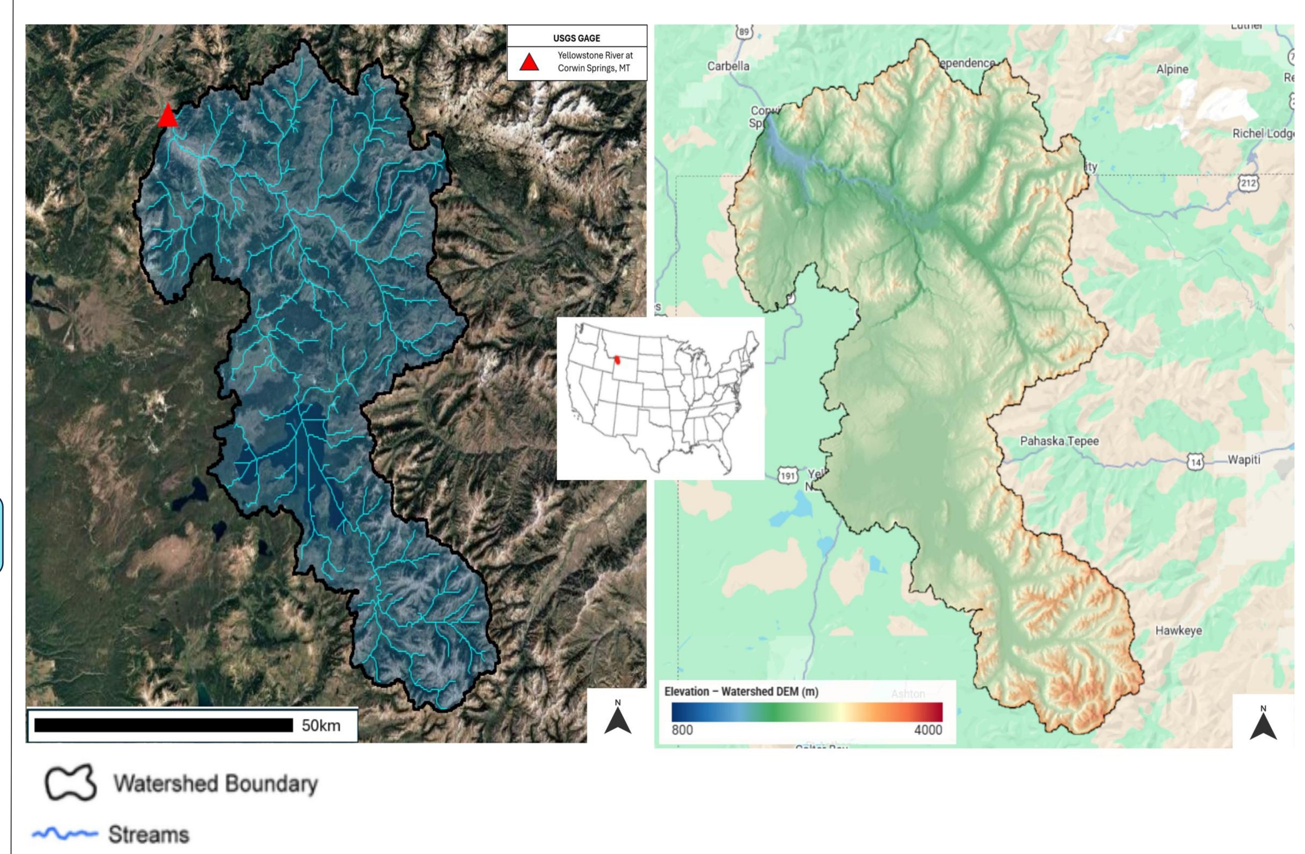



*Fig 1: Architecture for CNN-LSTM model for streamflow prediction*

- Daily streamflow, precipitation, mean temperature, and SWE are standardized and arranged into 30-day sliding windows to capture recent hydrologic and snow–climate memory for sub-seasonal forecasting.
- Separate models are trained for 7-, 14-, and 30-day lead times using data from 1981–2010, validated on 2011–2015, and tested on 2016–2023 with horizon-specific sequence generation, early stopping, and learning-rate scheduling to ensure stable training and fair comparison.
- The CNN-LSTM architecture uses a causal 1-D CNN for short-term feature extraction followed by stacked LSTMs to capture longer-term snowmelt-driven streamflow dependencies, with dropout and batch normalization improving generalization.
- For comparison, LSTM models rely on recurrence, Transformers use self-attention, and CNN-Transformer hybrids combine local feature extraction with global attention, with performance differences reflecting each model's ability to retain hydrologic memory at longer lead times.

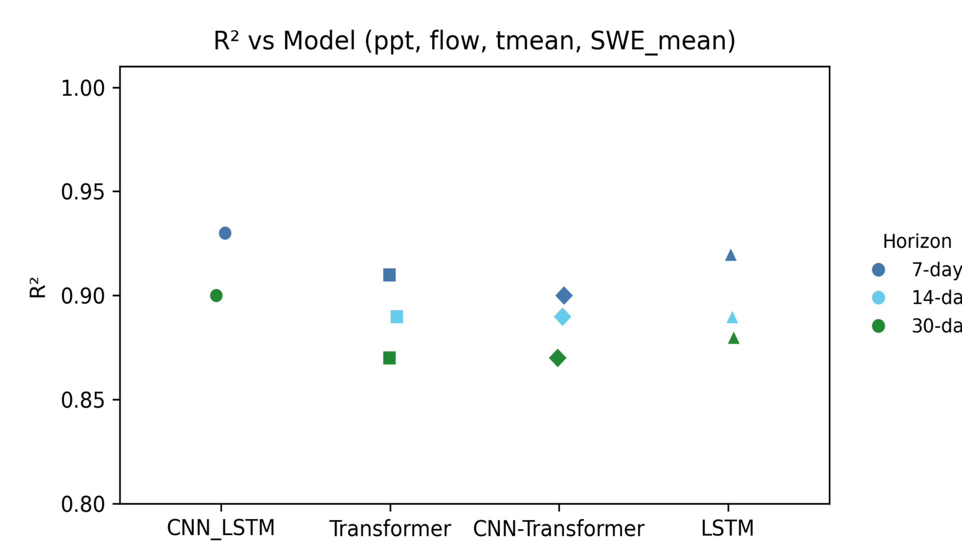
## STUDY AREA: YELLOWSTONE RIVER WATERSHED




- The Yellowstone River Basin, Montana, USA, is a snow-dominated mountainous watershed (~70,000 km²) with elevations ranging from ~1,200 to >4,000 m, where seasonal snow accumulation and spring snowmelt strongly control streamflow timing and magnitude. It is dominated by forested and alpine land cover in the headwaters, transitioning to grasslands, shrublands, and agricultural areas in the lower elevations downstream.

*Fig 2: Study Area: Yellowstone River Basin: Study Area, Stream Network and Elevation*

## RESULT 1: COMPARISON OF DEEP LEARNING MODEL PERFORMANCES




| Model | Forecast Horizons | RMSE | MAE | $R^2$ |
|---|---|---|---|---|
| LSTM | 7 | 33.51 | 15.25 | 0.92 |
| | 14 | 38.61 | 18.71 | 0.89 |
| | 30 | 39.3 | 18.66 | 0.85 |
| CNN-LSTM | 7 | 33.08 | 15.53 | 0.92 |
| | 14 | 38.25 | 17.91 | 0.89 |
| | 30 | 39.1 | 18.6 | 0.89 |
| Transformer | 7 | 36.88 | 18.39 | 0.9 |
| | 14 | 42.18 | 22.33 | 0.88 |
| | 30 | 38.64 | 21.35 | 0.87 |
| CNN-Transformer | 7 | 35.89 | 17.3 | 0.91 |
| | 14 | 38.3 | 18.76 | 0.85 |
| | 30 | 46.55 | 30.01 | 0.81 |

*Fig 3: Comparison of deep learning model performance ($R^2$) across different architectures evaluated using identical input features*

- The CNN–LSTM model shows consistently higher $R^2$ than other architectures when trained & validated with the same input features.
- This figure illustrates how different deep learning architectures perform across increasing forecast horizons under the same input conditions.

## RESULT 2a: PEAK TIMING VS MAGNITUDE PERFORMANCES

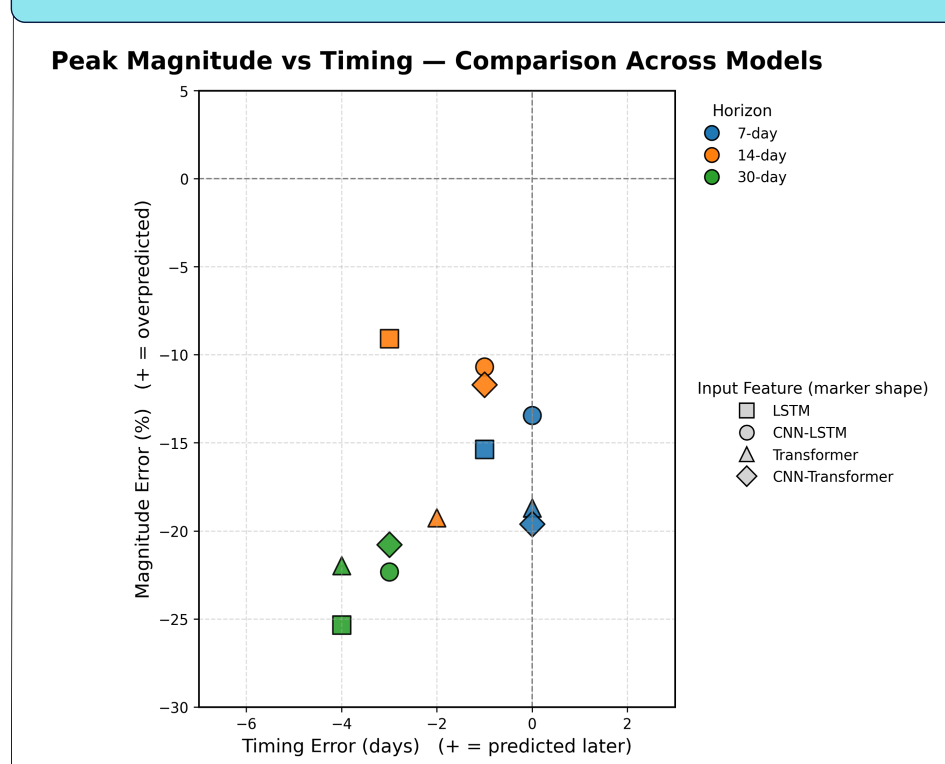



*Fig 4: Trade-off between peak streamflow magnitude error and peak timing error across deep learning models and forecast horizons*

- The figure relates peak timing error and peak magnitude error for 7, 14, and 30-day forecasts across deep learning models, illustrating trade-offs in peak prediction accuracy.
- With increasing lead time, models show larger peak underestimation and earlier predicted peaks, indicating growing uncertainty in extreme flow prediction at longer horizons.
- CNN-LSTM performs best overall, with errors clustering closest to zero, particularly for 7 and 14-day forecasts, demonstrating strong skill in capturing both peak timing and magnitude.
- Transformer-based models exhibit greater peak damping and earlier peak timing at 30-day lead times, reflecting reduced ability to represent snowmelt-driven extremes.

## RESULT 2b: PEAK FLOW PREDICTION BIAS ACROSS ALL MODELS

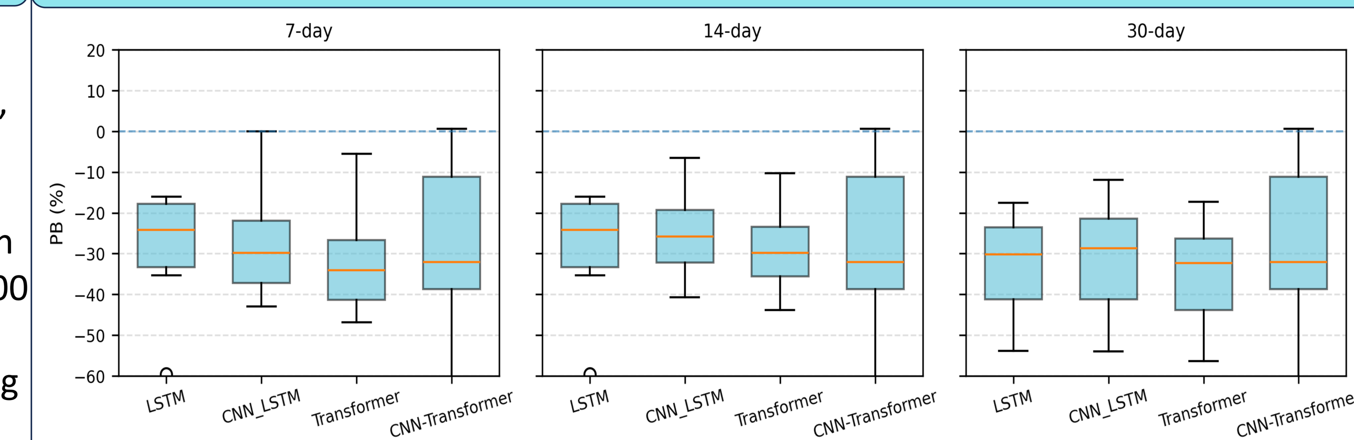
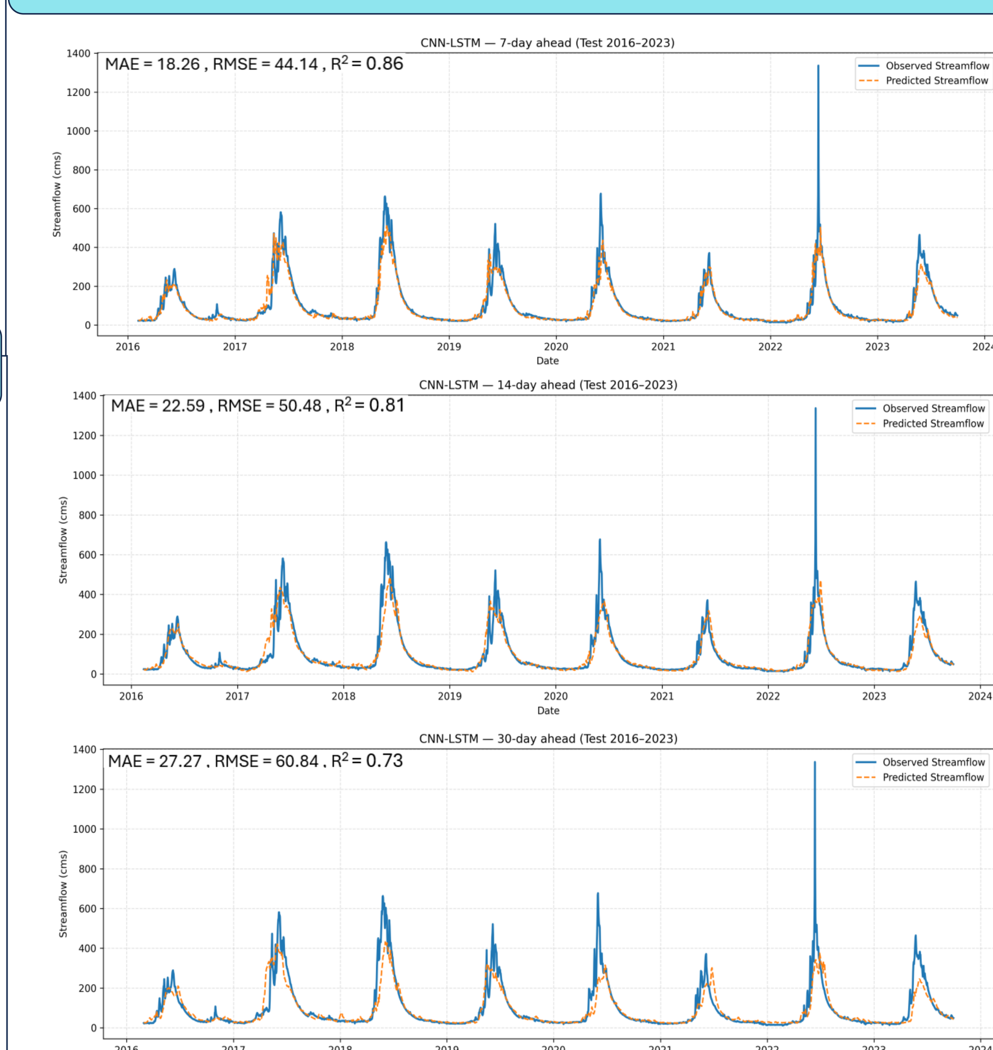



*Fig 5: Distribution of peak flow prediction bias across different deep learning architectures*

- The figure summarizes peak flow prediction bias for 7-, 14-, and 30-day forecasts across multiple deep learning architectures during the test period.
- All models increasingly underestimate peak flows at longer lead times, while CNN-LSTM and Transformer exhibit lower bias and more stable performance in capturing extreme streamflow events.

## RESULT 3: EFFECT OF SWE ON CNN-LSTM SUB-SEASONAL STREAMFLOW PREDICTION



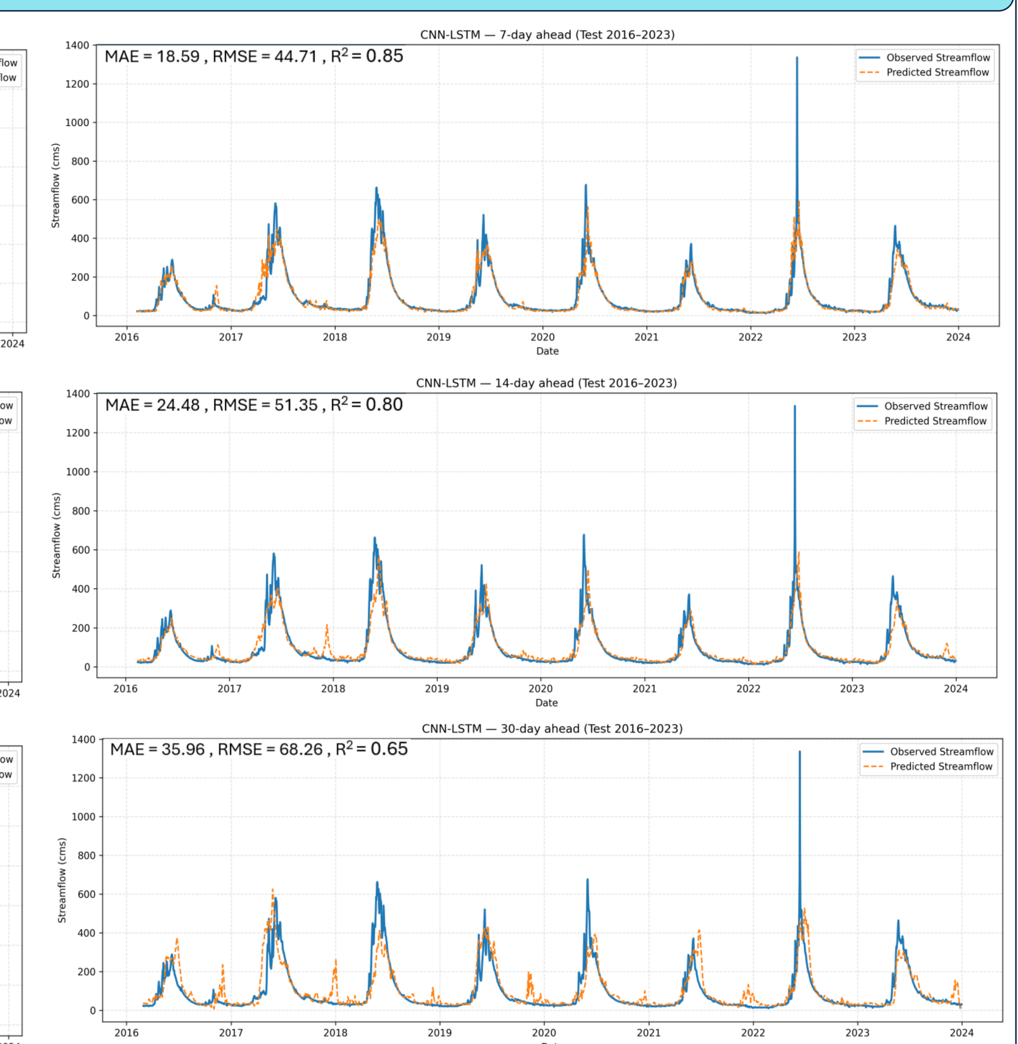




*Fig 6: Observed vs predicted streamflow with SWE* *Fig 7: Observed vs Predicted streamflow without SWE*

- Including SWE improves CNN-LSTM performance across forecast horizons, with lower RMSE and higher $R^2$ than the no-SWE case; the improvement is modest at 14 days (RMSE: 50.48 with SWE vs. 51.35 without SWE) but becomes more substantial at 30 days (RMSE: 60.84 with SWE vs. 68.26 without SWE), highlighting the increasing value of snow information at longer lead times.
- Without SWE, CNN-LSTM skill degrades more rapidly with increasing lead time, reflected by higher RMSE and lower $R^2$; while 7 and 14-day forecasts show similar $R^2$ (difference ≈ 0.01), the gap widens at 30 days ($R^2$ = 0.73 with SWE vs. 0.65 without SWE), indicating loss of hydrologic memory in longer-range forecasts.
- SWE-informed forecasts better preserve peak magnitude and timing, while models without SWE exhibit stronger peak damping and underestimation of extreme snowmelt-driven flows.
- The divergence between with and without-SWE results increases at longer lead times, highlighting SWE as a critical predictor for sustaining sub-seasonal forecast skill in snow-dominated basins.

## FUTURE WORK

- Apply the modeling framework to additional snow-dominated basins to assess the robustness and transferability if snow-informed deep learning models.
- Incorporate additional snow-related and climate variables to further improve representation of snow-driven hydrologic processes.
- Explore multi-horizon or multi-task learning frameworks to jointly predict streamflow at multiple lead times within a single model.

## ACKNOWLEDGEMENT

- This research was supported by NASA's Sub seasonal-to-Seasonal Hydrometeorological Prediction program (Grant No. 80NSSC24K1278), with feedback from student members of the Cho Lab at Texas State University.

## REFERENCES

- Koster, R. D., Mahanama, S. P., Livneh, B., Lettenmaier, D. P., & Reichle, R. H. (2010). Skill in streamflow forecasts derived from large-scale estimates of soil moisture and snow. *Nature Geoscience*, *3*(9), 613-616.
- Ng, K. W., Huang, Y. F., Koo, C. H., Chong, K. L., El-Shafie, A., & Ahmed, A. N. (2023). A review of hybrid deep learning applications for streamflow forecasting. *Journal of Hydrology*, *625*, 130141.