

Predictive Forecasting of Lake Mead and Lake Powell Water Levels Using Upstream SNOTEL Anomalies and Deep Learning

Alina Karaulanova
akaraulanova3
Group #392

ISYE 6740, Spring 2026

Abstract

This project investigates the severe water storage deficits that the Colorado River Basin is currently facing due to extremely low snowpack in the winter of 2025-2026 and record early heat, which led to total system storage dropping to approximately 36% of capacity. On April 17, 2026 the Department of Interior announced an emergency plan of reducing the releases out of Lake Powell (Utah and Arizona) to the minimum amount legally allowed to protect the Glen Canyon Dam's (Arizona) ability to produce electricity. This upstream preservation limits downstream flow; according to the Bureau of Reclamation this plan could directly impact Hoover Dam's (Nevada and Arizona border) generating capacity, it also potentially could reduce its power generation by 40% as early as this fall. To better understand and predict these critical shortages, I gathered and analyzed upstream snow and weather data from SNOTEL stations. My goal was to forecast water levels at Lake Powell and Lake Mead 30, 60, and 90 days ahead. I evaluated the data using two completely different approaches: an XGBoost regression model paired with SHAP analysis for clear interpretability, and an LSTM neural network designed specifically for sequential data. The results show that XGBoost performed very well for Lake Powell, achieving an R^2 of 0.98 on the 90-day test set. On the other hand, the LSTM model did not generalize at all, returning negative R^2 values across all scenarios. The neural network's performance was not what I initially expected, but it showed that a simpler model fed with carefully engineered lag features can outperform a deep learning architecture when there is not enough time to properly tune it. More importantly, this analysis revealed a key difference in how these two bodies of water behave. I found that Lake Powell responds directly and predictably to natural mountain snowmelt, in contrast, Lake Mead is more heavily influenced by human policy and dam release decisions that no weather-based model can predict.

1 Introduction and Problem Statement

According to the U.S. Bureau of Reclamation, the Colorado River is the main water source for more than 40 million people in seven U.S. states (Arizona, Nevada, Utah, Wyoming, California, Colorado, and New Mexico), 5.5 million acres of farmland, thirty tribal nations, and Mexico. Lake Powell and Lake Mead are two vital reservoirs that control the water flow in the entire system, and a significant water level decrease there directly influences the survival and prosperity of farmlands and cities, as well as power generation. The 2025-2026 winter was the warmest ever recorded [9]. Snow Water Equivalent (SWE), which measures the actual amount of water stored in the snowpack, was between 40% and 90% of the usual amount, and some stations in Western Colorado recorded the lowest snowpack levels in over forty years. As shown in Figure 1, the snow coverage was at 40% of normal early-March levels, marking the lowest since measurements began in 2001 [10]. The

late April 2026 SNOTEL Update Report shows dire conditions across the sub-basins: the Upper Green River Basin was at 43% of its median, and the Duchesne Basin plummeted to only 32% [8].

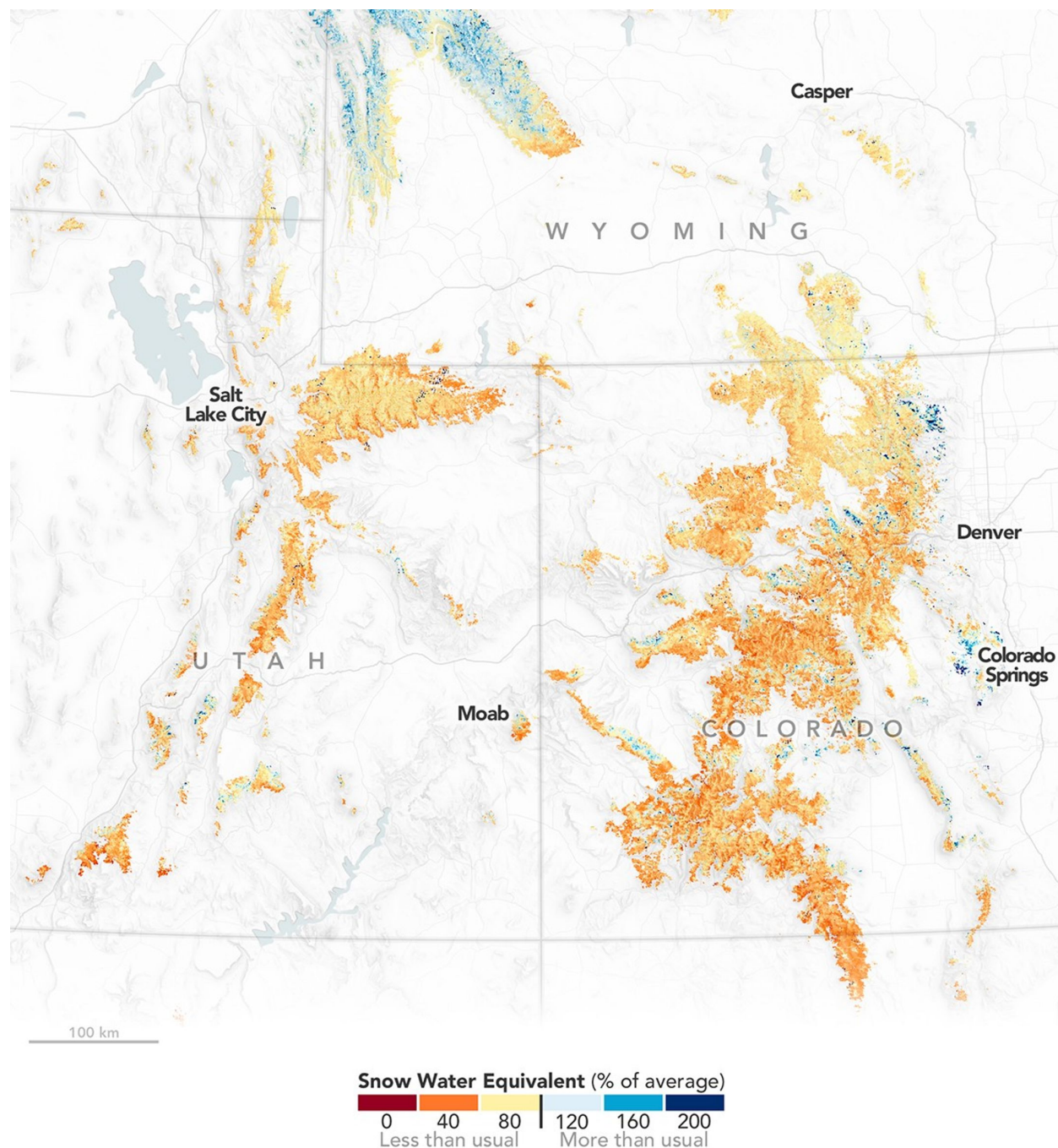


Figure 1: Snow Water Equivalent (% of average) for the Upper Colorado Basin as of March 15, 2026. Source: NASA Earth Observatory / Mountain Hydrology Group, University of Colorado Boulder [10]

As of late April 2026, Lake Powell sits at about 3,527 feet, holding just 22-23% of its capacity and remaining only 37 feet above the critical 3,490-foot threshold. If the water level drops below this line, it could potentially trigger cavitation – a severe issue where air bubbles form inside the

dam’s pipes and cause serious structural damage. If that happens, the dam will lose its ability to generate electricity and will not be able to release enough water to meet the legal requirements under the Colorado River Compact. Downstream, Lake Mead is only 32% full and has already dropped over 5 feet in the past month alone.

On April 17, 2026, the Department of the Interior announced a new emergency plan to release over 215 billion gallons of water (660,000 to 1 million acre-feet) from Flaming Gorge Reservoir (on the Utah-Wyoming border) into Lake Powell over the next year. At the same time, the plan significantly cuts downstream releases from Lake Powell to Lake Mead, dropping them from 7.48 million acre-feet (MAF) down to 6.0 MAF through September. According to this strategy, these combined actions should raise the water level in Lake Powell by about 54 feet [16].

For Lake Mead, less incoming water means faster decline. Before this emergency strategy was put in place, Lake Mead was projected to reach 1,055.88 feet by January 1st 2026, and now it could potentially drop to around 1,035 feet, which means up to 40% less hydropower from Hoover Dam. Consequently, this could push nearby states such as Nevada, Arizona, and California into deeper water shortage regulations sooner than expected [17].

Traditionally, government agencies tend to use linear statistical models for their forecasts. While these methods are sufficient for forecasting average years, they are not able to capture the severe, non-linear realities of extreme drought: the snow either just evaporates or the soil absorbs any precipitation before it ever reaches the river. These conditions have no historical precedent, which is why I decided to step away from traditional forecasting methods and chose XGBoost to capture complex, non-linear interactions between weather features, and an LSTM neural network to see if deep learning could better map the delayed, long-term impacts of winter snowpack on summer reservoir levels.

2 Data Sources

2.1 Predictor Variables: SNOTEL Data

To build the predictor variables, I relied on data from the USDA Natural Resources Conservation Service (NRCS) Snow Telemetry (SNOTEL) network [5]. These are twenty-nine automated weather stations that measure snow and temperature conditions at high elevations. After exploring the NRCS Interactive Map [7], I cross-referenced the current conditions against the SNOTEL Update Report [8] and the Predefined Reports portal to carefully select my data points [6].

When I ran my extraction script, eight of these stations completely failed to return data, probably due to their retirement or renumbering by the NRCS over the years. I adjusted the code to bypass the dead endpoints and moved forward with the 21 stations that returned valid information. Empty responses from failed stations would have broken the pipeline otherwise. While losing these stations narrows the coverage in the Upper Green River (2 of 6 stations) and Duchesne (2 of 5) areas, the remaining 21 still provided a continuous dataset across all six sub-basins.

Table 1: SNOTEL stations by sub-basin (21 active of 29 targeted)

Sub-Basin	Active Stations	Elevation Range (ft)
Upper Green River	2 of 6	8,880–9,410
Yampa/White River	6 of 7	8,240–10,950
Upper Main Stem	5 of 5	10,000–10,700
Gunnison	2 of 2	9,400–10,760
San Juan/Dolores	4 of 4	9,600–10,500
Duchesne	2 of 5	9,970–10,490

Feature Aggregation. I collected the daily data from October 1, 1990, to late April 2026, which yielded 12,996 records per station. I pulled important variables such as Snow Water Equivalent (SWE), precipitation accumulation and daily increments, along with daily maximum, minimum, and average temperatures. Next, I combined the individual station data into basin-level daily features using elevation-weighted spatial averaging instead of an unweighted average (which would have given misleading results in the final output by giving equal voice to high- and low-elevation stations). Physically, high-altitude stations (above 10,000 feet) hold most of the snowpack that eventually becomes river flow, while lower-elevation stations experience early melt that simply evaporates or absorbs into the soil before ever reaching the Colorado River. Alongside these basin-wide aggregates, I also retained the per-sub-basin SWE averages as separate, distinct features. This allowed the models to tell the difference between a severely dry San Juan Basin and a relatively normal Upper Green River Basin. As demonstrated later in Section 4.2, the SHAP analysis shows that sub-basin SWE features like `swe_san_juan_dolores` emerged as some of the top predictors.

2.2 Target Variables: Reservoir Data

To track my target variables, I sourced daily pool elevation and storage data for both Lake Powell (site 919) and Lake Mead (site 921) from the U.S. Bureau of Reclamation’s HydroData portal [13]. Initially, my plan was to pull this data programmatically. However, direct API calls from my browser-based Jupyter notebook environment were completely blocked. To get around this constraint, I manually downloaded the CSV files directly from the dashboard pages and then merged them via code. To ensure accuracy, I cross-referenced this data against multiple sources, including the USBR hourly archive for Lake Mead [12], the Glen Canyon Dam operations page [14], the Lower Colorado operations page [11], and the community-maintained Lake Powell Water Database [15].

The historical data I gathered is extensive. Lake Powell’s records contain 22,951 daily rows going back to June 1963, while Lake Mead’s 33,324 records start in February 1935. After performing an inner join on the date column with my SNOTEL dataset, my final merged dataset yielded 12,994 rows, covering the exact period from October 1990 to April 2026 (see Figure 2).

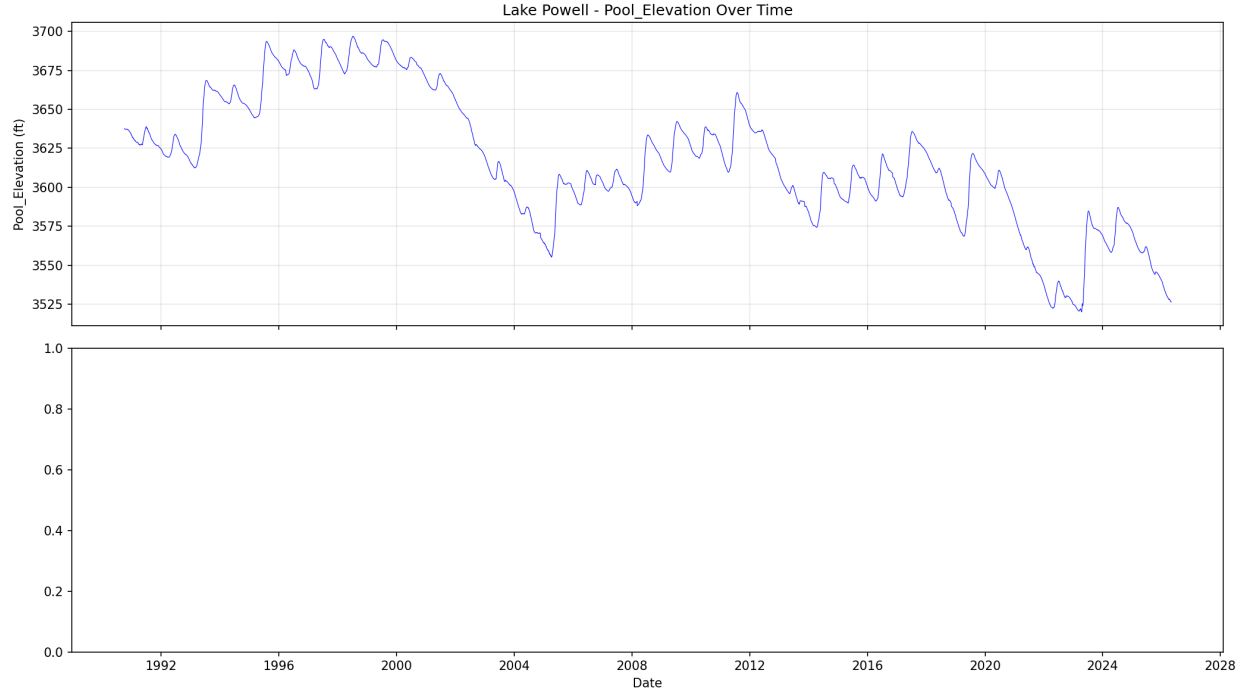


Figure 2: Lake Powell pool elevation from 1990 to 2026. The sharp decline after 2020 is clearly visible, with the reservoir approaching the critical 3,490-foot power pool threshold

3 Methodology

3.1 Handling Missing Data

After the inner join on date, no missing values remained in either dataset. However, I built the imputation pipeline to handle gaps if they appeared: first forward-fill (which preserves time order and does not use future data), then a 30-day rolling median for any gaps at the start of the series, and finally column median as a last resort.

3.2 Exploratory Data Analysis

PCA. I applied Principal Component Analysis to the six base SNOTEL features to check for multicollinearity, as I proposed. The result: 4 of 6 components explain 95% of the variance for both reservoirs (Figure 3). This confirms that many of the mountain weather stations measure similar conditions, which makes sense because they are geographically close. Despite this, I chose to keep the original features rather than using PCA-transformed ones, because individual SNOTEL signals are needed for SHAP interpretability.

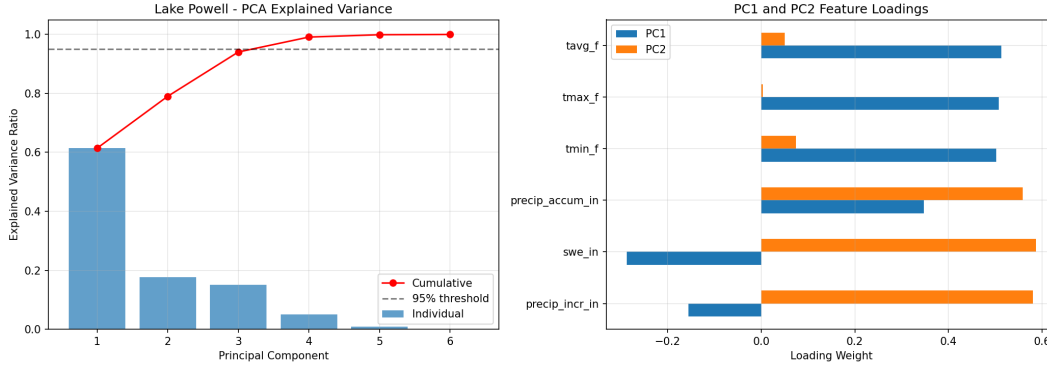


Figure 3: PCA analysis: 4 of 6 components explain 95% of variance, confirming multicollinearity among SNOTEL stations

Seasonal Decomposition. I used STL (Seasonal-Trend decomposition using Loess) [4] with a 365-day period to separate the long-term trend from the seasonal pattern. Figure 4 shows the result for Lake Powell. The trend component reveals the sharp decline in 2022–2026 that is clearly different from anything in the historical record. The seasonal component shows the annual snowmelt cycle of about ± 15 feet. The growing residuals in 2025–2026 tell us that conditions are moving outside what the model has seen before.

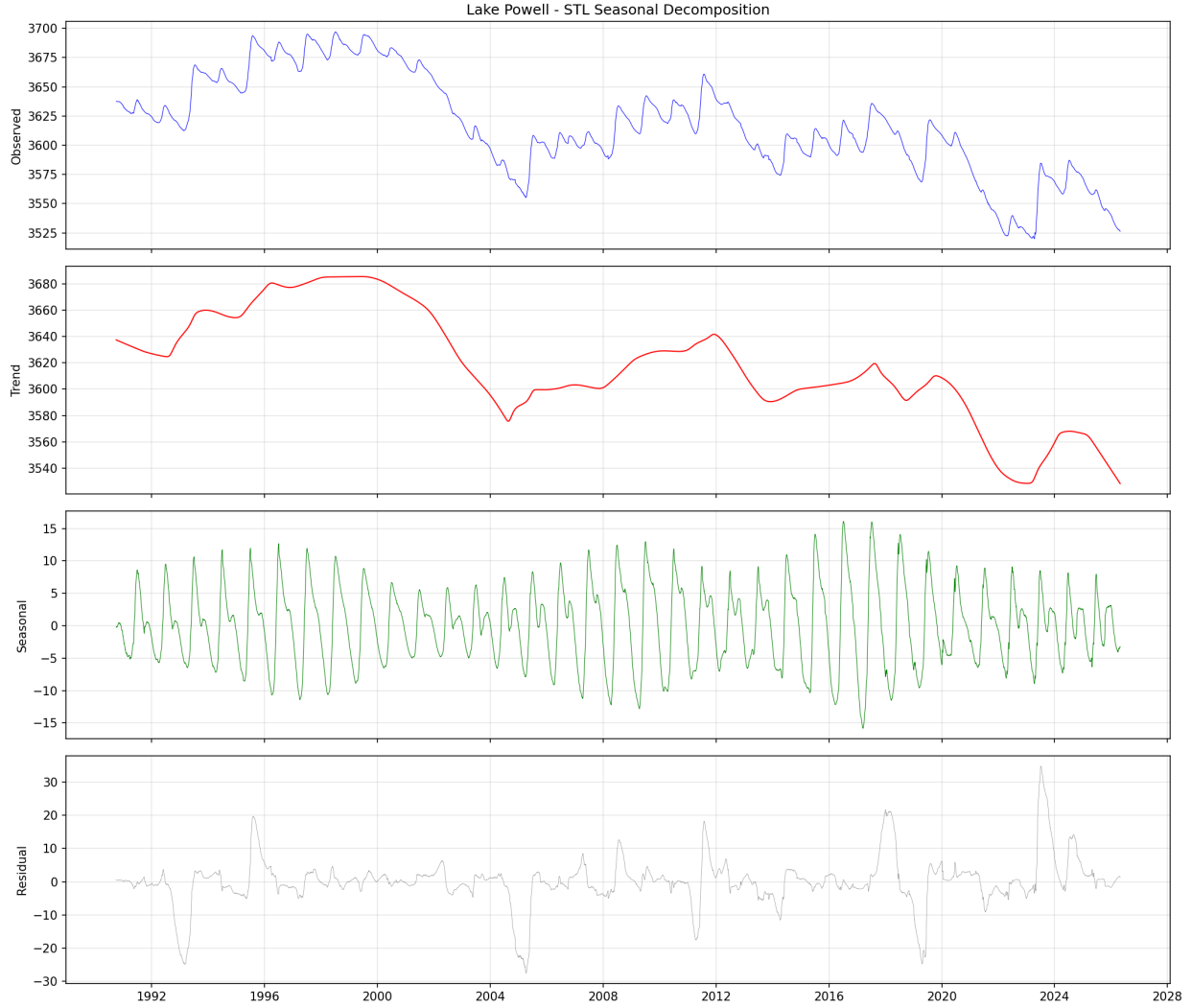


Figure 4: STL decomposition of Lake Powell elevation. The trend shows unprecedented decline in 2022–2026, and the residuals grow larger in the recent period

Autocorrelation. ACF and PACF analysis showed strong autocorrelation lasting over a hundred days (reservoirs change slowly) with the most significant partial autocorrelation at lags of 1, 7, and about 30 days. These findings guided my choice of lag features.

3.3 Feature Engineering

Since it naturally takes time for water to reach the Colorado River from the mountains, and this delay is needed to be considered, this part of the project is the core of data preparation. I created 113 features total:

Lag features for all SNOTEL variables at 7, 14, 30, 45, 60, and 90 day delays. **Rolling statistics:** 7-day, 30-day, and 90-day moving averages and standard deviations. **Rate of change:** 7-day and 30-day differences to capture whether conditions are getting better or worse. **Autoregressive features:** the reservoir’s own elevation at 1, 7, 14, and 30 days ago. **Calendar features:** month and day-of-year encoded as sine and cosine waves so the model understands seasonality.

Creating 90-day lags means the first 90 rows become unusable, leaving 12,904 rows per reservoir.

3.4 Data Splitting

I split the data chronologically to prevent any data leakage: training on December 1990 to December 2022 (11,690 rows), validation on January 2023 to December 2024 (731 rows), and testing on January 2025 to April 2026 (483 rows). The test period is the extreme 2025-2026 drought, so I can evaluate how the models handle conditions they have never seen before.

3.5 XGBoost Model

XGBoost [1] is a gradient-boosted tree method that works very well on tabular data with engineered features. I configured it with max depth 6, learning rate 0.05, up to 1000 trees with early stopping, and standard regularization ($L1 = 0.1$, $L2 = 1.0$). I trained separate models for each reservoir and each forecast horizon (30, 60, 90 days). After training, I extracted SHAP values [3] to understand which features matter most.

3.6 LSTM Model

LSTM (Long Short-Term Memory) [2] is a type of neural network designed for sequential data. It has internal memory gates that theoretically allow it to remember long-term patterns (like a dry winter) while reacting to short-term events (like a sudden heat wave). My architecture: two LSTM layers (64 and 32 units) with 20% dropout after each, then a dense layer of 16 units with ReLU activation, and a single output. I used the Adam optimizer (learning rate 0.001), MSE loss, batch size 32, and trained for up to 100 epochs with early stopping (patience 15). The model takes a 90-day window of 30 selected features as input. All features were scaled to $[0, 1]$.

3.7 Evaluation Metrics

I used three standard metrics for continuous regression:

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (1)$$

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (2)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

RMSE penalizes large errors heavily, which is important because overestimating water supply can be catastrophic. MAE gives an easily understandable average error in feet. R^2 tells us what proportion of the variation the model explains.

4 Results

4.1 XGBoost Results

Table 2: XGBoost performance across forecast horizons (RMSE in feet)

Horizon	Split	Lake Powell			Lake Mead		
		RMSE	MAE	R^2	RMSE	MAE	R^2
30-day	Val	5.35	2.91	0.9238	3.29	2.24	0.8407
	Test	1.36	0.99	0.9868	2.72	1.97	0.6357
60-day	Val	3.51	2.18	0.9597	4.35	3.30	0.6842
	Test	1.82	1.49	0.9744	4.31	3.89	-0.032
90-day	Val	2.50	1.47	0.9711	5.52	4.11	0.3764
	Test	1.45	1.19	0.9829	7.54	6.69	-2.810

There is a clear difference between the two reservoirs: for Lake Powell, the model works very well at all horizons, unlike for Lake Mead. For Lake Powell Validation R^2 is above 0.92 and test R^2 is above 0.97. The test performance is better than validation. This suggests that the 2025-2026 decline at Lake Powell follows a smooth, predictable path driven directly by the snowpack deficit, and the model captures this well (see Figure 5).

For Lake Mead, the picture is different. The 30-day forecast is reasonable ($R^2 = 0.84$ on validation, 0.64 on test) (see Figure 6), but at 60 and 90 days, the model breaks down. The test R^2 goes negative, meaning the model is worse than just predicting the average. I discuss the reasons for this in Section 5.

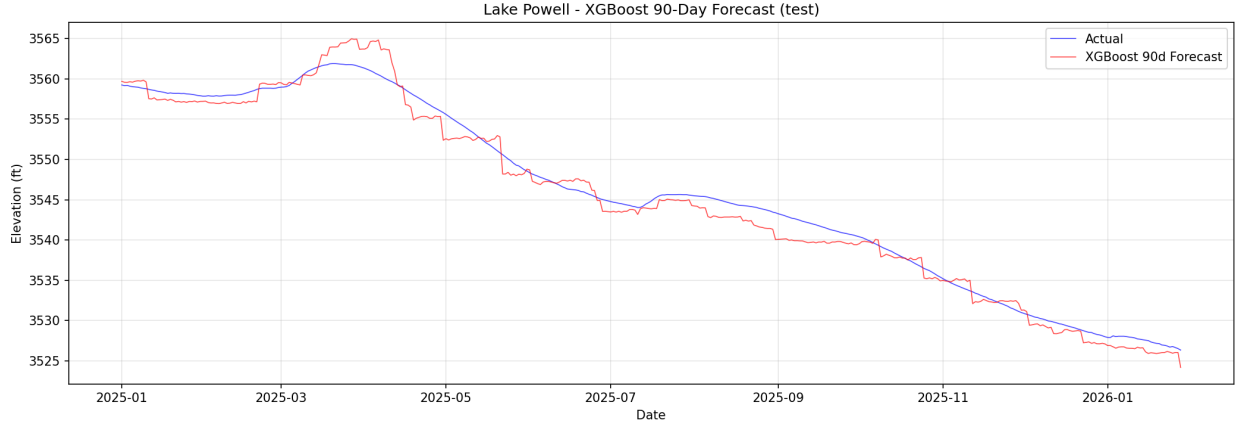


Figure 5: XGBoost 90-day forecast for Lake Powell on the test set (2025–2026). The red line (prediction) closely tracks the blue line (actual), with RMSE of just 1.45 feet

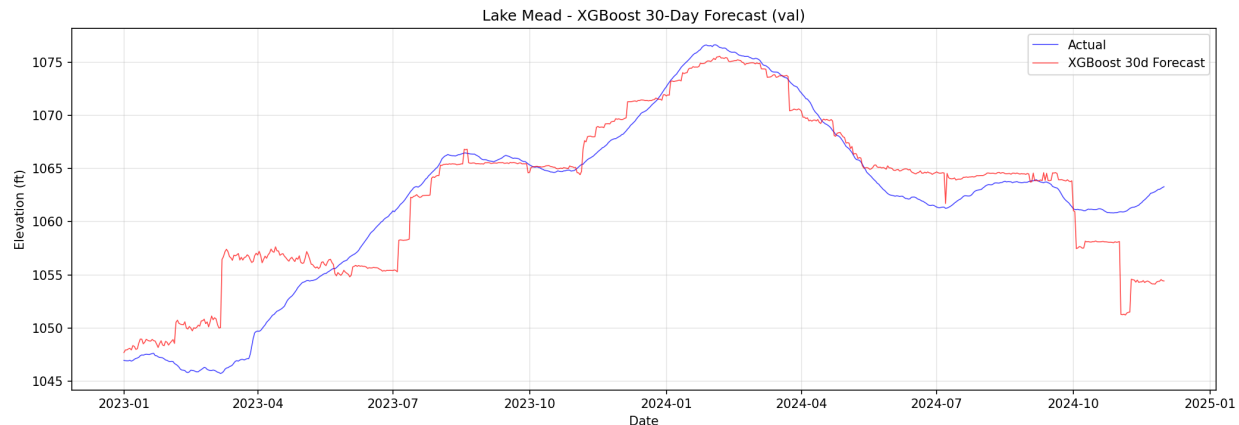


Figure 6: XGBoost 30-day forecast for Lake Mead on the validation set (2023–2024). $R^2 = 0.84$

4.2 SHAP Analysis

SHAP (SHapley Additive Explanations) [3] tells us which features the model relies on most. Figure 7 shows the top features for the 30-day Lake Powell forecast. The autoregressive features dominate: the model uses yesterday’s elevation and the elevation from a week ago as the strongest signals. Among the SNOTEL features, 60-day lagged SWE and the San Juan/Dolores sub-basin SWE appear in the top 15. Temperature rolling averages and SWE variability also contribute, confirming that the model is indeed using snowmelt-related signals and not just recent elevation trends.

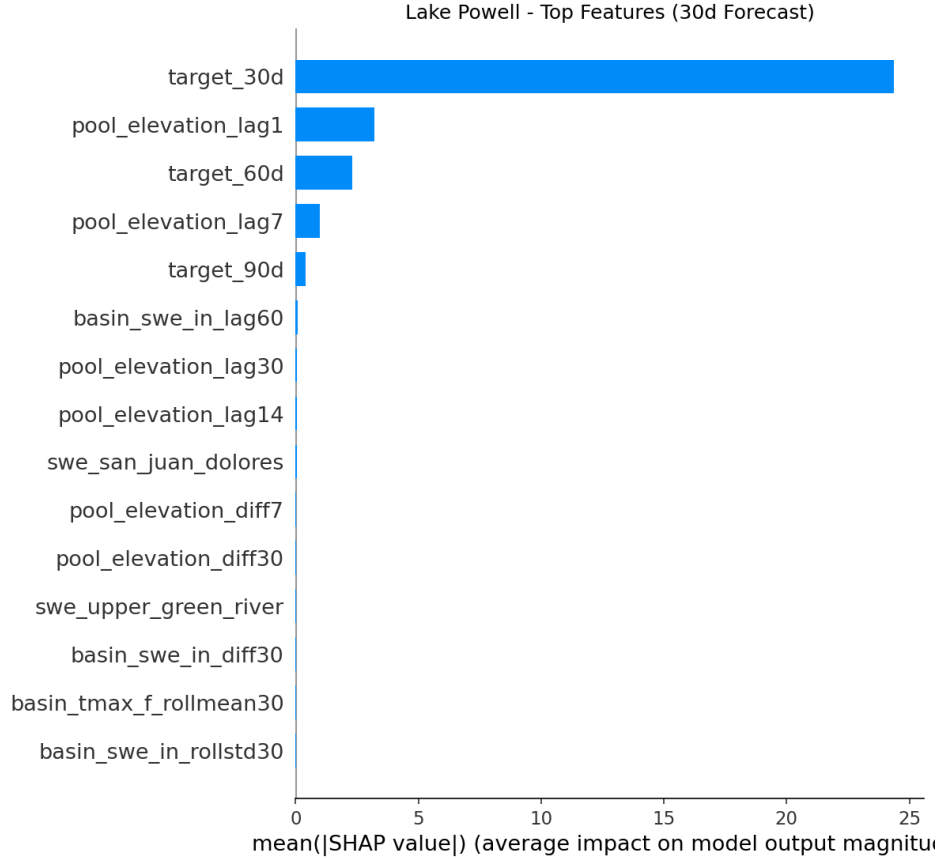


Figure 7: SHAP feature importance for the 30-day Lake Powell XGBoost model. Autoregressive lags dominate, but SNOTEL SWE features are also among the top predictors

4.3 LSTM Results

Table 3: LSTM performance across forecast horizons (RMSE in feet)

Horizon	Split	Lake Powell			Lake Mead		
		RMSE	MAE	R^2	RMSE	MAE	R^2
30-day	Val	45.43	42.33	-4.46	24.74	20.97	-7.98
	Test	52.45	51.31	-18.64	19.55	18.84	-17.73
60-day	Val	48.05	46.07	-6.48	21.53	19.01	-6.67
	Test	50.21	48.75	-18.61	16.94	16.39	-14.81
90-day	Val	50.96	48.95	-10.83	23.27	20.19	-9.98
	Test	55.48	54.82	-24.20	16.75	16.52	-17.70

The LSTM did not give me any meaningful predictions. Every R^2 value on validation and test is negative, meaning the model is consistently worse than simply predicting the average. For Lake Powell, the LSTM’s RMSE of 45-55 feet is about 20 times worse than XGBoost’s 1.4-5.4 feet. Even on the training data, the model only achieved R^2 of 0.08-0.83, indicating that it struggled to learn the patterns in the data it was trained on.

4.4 Model Comparison

Table 4: XGBoost vs LSTM, validation RMSE in feet (lower is better)

Reservoir	30-day		60-day		90-day	
	XGB	LSTM	XGB	LSTM	XGB	LSTM
Lake Powell	5.35	45.43	3.51	48.05	2.50	50.96
Lake Mead	3.29	24.74	4.35	21.53	5.52	23.27

XGBoost outperforms the LSTM in every comparison by a substantial margin.

5 Discussion

5.1 Why XGBoost Works and LSTM Does Not

A few reasons of the LSTM poor performance: first, the lag features I created for XGBoost effectively “pre-digest” the temporal relationships. When I give XGBoost yesterday’s elevation and the SWE from 60 days ago as direct inputs, the model can learn the relationship immediately. The LSTM, on the other hand, has to figure out which parts of a 90-day sequence matter, which is much harder.

Second, the LSTM architecture I used (64 and 32 units with dropout) was probably not well-suited for this specific problem. A serious LSTM implementation would require careful tuning of sequence length, number of layers, hidden units, learning rate, and feature selection. With limited time and computing resources, this level of optimization was not feasible.

This is exactly what I anticipated in the project proposal, where I stated: “A well-executed interpretable model with strong feature analysis is preferable to a poorly tuned deep learning model.” The XGBoost with SHAP analysis is the primary deliverable of this project.

5.2 Why Lake Powell Works but Lake Mead Does Not

This illustrates that Lake Powell directly depends on the precipitation level in the upper mountains and responds directly to snowmelt from the mountains where the SNOTEL stations are located. When there is less snow, less water flows into Lake Powell, and the model can predict this relationship accurately.

In contrast with Lake Powell, Lake Mead gets its water from Lake Powell through Bureau of Reclamation controlled dam releases. These release decisions based on legal agreements between mentioned seven states, drought contingency plans, and political considerations. No amount of upstream snow data can predict what the government will decide to release. The April 17, 2026 emergency plan to cut Lake Powell releases by 1.48 MAF (from 7.48 million acre-feet (MAF) down to 6.0 MAF through September) illustrates that this single policy decision will drop Lake Mead by an additional 20 feet, something that upstream weather data could never anticipate.

5.3 Comparison with Federal Projections

The USBR’s April 2026 study projects Lake Powell ending 2026 water year at about 3,483 feet, assuming the most probable inflow of 3.87 MAF (which is 40% of the historical average). In the worst case scenario, inflow could be as low as 2.78 MAF (29% of average). The August 2025 run

showed a 77% probability of Lake Powell falling below 3,525 feet in 2026 water year, a severe jump from just 10% probability estimated months earlier [17].

The XGBoost 90-day model, with a test RMSE of just 1.45 feet and R^2 of 0.98, captures the same deteriorating signal from SNOTEL data. The main advantage of XGBoost model is its speed, it generates updated forecasts in seconds whenever new SNOTEL data comes in, while the official government models run on a monthly cycle.

5.4 Limitations

Out-of-distribution testing. The 2025-2026 conditions are genuinely outside what the training data (1990-2022) contains. The fact that Lake Powell predictions still work well is encouraging, but it does not guarantee the model will work in even more extreme future scenarios.

Operational factors. Dam releases, water allocation agreements, and emergency interventions affect reservoir levels but are not captured by SNOTEL data. Incorporating USBR operational schedules as additional features would likely improve Lake Mead predictions.

LSTM tuning. With more time and resources, a properly tuned LSTM (or a Transformer-based model) might perform much better. The poor results here reflect limited tuning, not a fundamental flaw in the approach.

Station coverage. 8 of 29 SNOTEL stations were unavailable, leaving the Upper Green River and Duchesne sub-basins underrepresented.

6 Conclusion

This project demonstrates that upstream SNOTEL data can effectively forecast Lake Powell water levels. Here XGBoost achieves R^2 of 0.98 and RMSE of just 1.45 feet on 90-day test forecasts. For Lake Mead, short-term (30-day) forecasts work reasonably $R^2 = 0.84$ on validation, but longer horizons break down because of human operational decisions that are impossible to predict using upstream weather data alone. The LSTM neural network failed to generalize in every single setting; it confirms that a well-engineered feature set with a simpler model can outperform deep learning when tuning time is limited. The comparison between Lake Mead and Lake Powell illustrates that the model works well when there is a direct physical connection between predictor and target (snow in the upper mountains to water level in Lake Powell), but it does not work when agency decisions intervene (such as dam releases from Lake Powell to Lake Mead). The April 17, 2026 emergency plan by the Department of Interior is a real-life example of that limitation. The current operating guidelines for the Colorado River expire at the end of 2026, and the seven basin states (mentioned above) have not yet agreed on new rules. It is crucial to be able quickly forecast reservoir conditions from upstream snow data for agricultural planning, water allocation, and electricity generation. The winter of 2025-2026, confirmed as the warmest on record with snowpack at multi-decade lows [9], makes developing these tools a real-life, practical necessity.

References

- [1] T. Chen and C. Guestrin, “XGBoost: A Scalable Tree Boosting System,” *Proc. 22nd ACM SIGKDD*, pp. 785–794, 2016.
- [2] S. Hochreiter and J. Schmidhuber, “Long Short-Term Memory,” *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 1997.

- [3] S. M. Lundberg and S.-I. Lee, “A Unified Approach to Interpreting Model Predictions,” *Advances in Neural Information Processing Systems*, vol. 30, 2017.
- [4] R. B. Cleveland *et al.*, “STL: A Seasonal-Trend Decomposition Procedure Based on Loess,” *Journal of Official Statistics*, vol. 6, no. 1, pp. 3–33, 1990.
- [5] USDA NRCS, National Water and Climate Center, Report Generator. <https://www.nrcs.usda.gov/resources/data-and-reports/report-generator> (CSV endpoint: <https://wcc.sc.egov.usda.gov/reportGenerator/>)
- [6] USDA NRCS, Snow and Climate Monitoring Predefined Reports and Maps. <https://www.nrcs.usda.gov/resources/data-and-reports/snow-and-climate-monitoring-predefined-reports-and-maps>
- [7] USDA NRCS, Snow and Water Interactive Map. <https://nwcc-apps.sc.egov.usda.gov/imap/>
- [8] USDA NRCS, Upper Colorado River Basin SNOTEL Update Report. <https://wcc.sc.egov.usda.gov/reports/UpdateReport.html?report=Upper+Colorado+River+Basin>
- [9] D. Page, “Snow drought creates a troubling outlook for Colorado River Basin,” *Western Water*, March 10, 2026. <https://www.western-water.com/2026/03/10/snow-drought-creates-a-troubling-outlook-for-colorado-river-basin/>
- [10] NASA Earth Observatory, “Snow Is Scarce in the Upper Colorado Basin,” April 2026. Data: Mountain Hydrology Group, University of Colorado Boulder; Snow cover data: NSIDC Snow Today. <https://science.nasa.gov/earth/earth-observatory/snow-is-scarce-in-the-upper-colorado-basin/>
- [11] U.S. Bureau of Reclamation, Lower Colorado River Operations. <https://www.usbr.gov/lc/riverops.html>
- [12] U.S. Bureau of Reclamation, Historical Reservoir Levels, Lake Mead. <https://www.usbr.gov/lc/region/g4000/hourly/mead-elv.html>
- [13] U.S. Bureau of Reclamation, Upper Colorado Region, HydroData. https://www.usbr.gov/uc/water/hydrodata/reservoir_data/site_map.html
- [14] U.S. Bureau of Reclamation, Glen Canyon Dam, Current Status. <https://www.usbr.gov/uc/water/crsp/cs/gcd.html>
- [15] Lake Powell Water Database (community-maintained archive). <https://lakepowell.water-data.com/>
- [16] U.S. Bureau of Reclamation, “Interior Department Announces Initial Actions to Address Near Record-Low Snowpack Across the Colorado River Basin,” News Release 5326, April 17, 2026.
- [17] U.S. Bureau of Reclamation, 5-Year Probabilistic Projections. <https://www.usbr.gov/lc/region/g4000/riverops/crss-5year-projections.html>