NM WRRI Student Water Research Grant Final Report

1. Student Researcher: Thanushan Kirupairaja, PhD student, MSC-3CE, Box 30001, New Mexico State University, Las Cruces, NM 88003, kthanu@nmsu.edu, (575) 646-6044; Civil Engineering, PhD in Civil Engineering-Water Resources

Faculty Advisor: Dr. A. Salim Bawazir, Civil Engineering Department, MSC-3CE, Box 30001, New Mexico State University, Las Cruces, NM 88003; abawazir@nmsu.edu; (575) 646-6044

2. Project title: Reservoir Evaporation Prediction Under Climate Change Scenarios Using Artificial Intelligence in Predictive Modeling of Lake Evaporation in Caballo Reservoir, New Mexico, USA.

3. Description of research problem and research objectives

Climate change observations, including recent global record-breaking temperatures, are leading to direct or indirect impacts on various environmental processes. Climate change continues to alter temperature patterns, precipitation, evaporation, and other meteorological variables. These changes will affect the volume of evaporation from reservoirs and lakes, impacting water availability for agriculture, municipal, and hydroelectric power generation. More advanced, accurate, and scalable methods to estimate and forecast reservoir and lake evaporation are necessary to operate reservoirs efficiently in lieu of climate change. Farmers in the south-central part of New Mexico, El Paso, Texas, and Juarez, Mexico, rely on surface water stored in Elephant Butte and Caballo reservoirs to irrigate their crops. Therefore, the management and operation of these reservoirs are essential to the agro-economy of the region and the nation.

Currently, a project led by Dr. Salim Bawazir at NMSU to measure evaporation losses at Elephant Butte and Caballo reservoirs and sponsored by the US Bureau of Reclamation is underway, and this proposed project will be a part of this larger project. Artificial Intelligence (AI) has shown promising capabilities in modeling complex systems and has been successfully applied in various environmental studies. However, its use in estimating and forecasting reservoir evaporation losses is limited, and this project tends to explore its capabilities, specifically in estimating and forecasting evaporation losses from the Caballo reservoir. The specific objectives of this proposed study are to **i)** develop an AI-driven model that integrates the bulk aerodynamic (B_AER) evaporation method for forecasting lake evaporation rates under climate change **ii)** validate the AI model's predictions against measured data. The long-term goal is to develop a tool that aids policymakers, environmentalists, and researchers in water resource management and climate change mitigation strategies.

4. Description of methodology employed

4.1 Site Description

The Caballo Reservoir (CBR), formed by the construction of Caballo Dam (N32° 53′ 48″, W107° 17′ 31″; WGS 84), is situated 25 miles downstream from Elephant Butte Reservoir. The dam was built between 1936 and 1938. It was primarily intended for flood control and to manage non-irrigation hydropower releases from Elephant Butte Dam. Caballo Dam is an earth-fill zoned structure stands 96 feet high with a crest length of 4,558 feet. According to a sedimentation survey conducted by Ferrari in 2007, the CBR has a total capacity of 324,934 acre-feet and surface area of 11,532 acres at a reservoir elevation of 4,182 feet (US Bureau of Reclamation projected vertical datum). At this elevation, the reservoir extends to about 17 miles long with an average width of 1 mile. From January 2021 to December 2023, the CBR's storage capacity varied from 6,049 to 58,055 acre-feet. During the irrigation season, when water is released from Elephant Butte Reservoir and stored in CBR, the reservoir's surface area extends northward from Caballo Dam. The location of the Caballo Reservoir is shown in Figure 1.
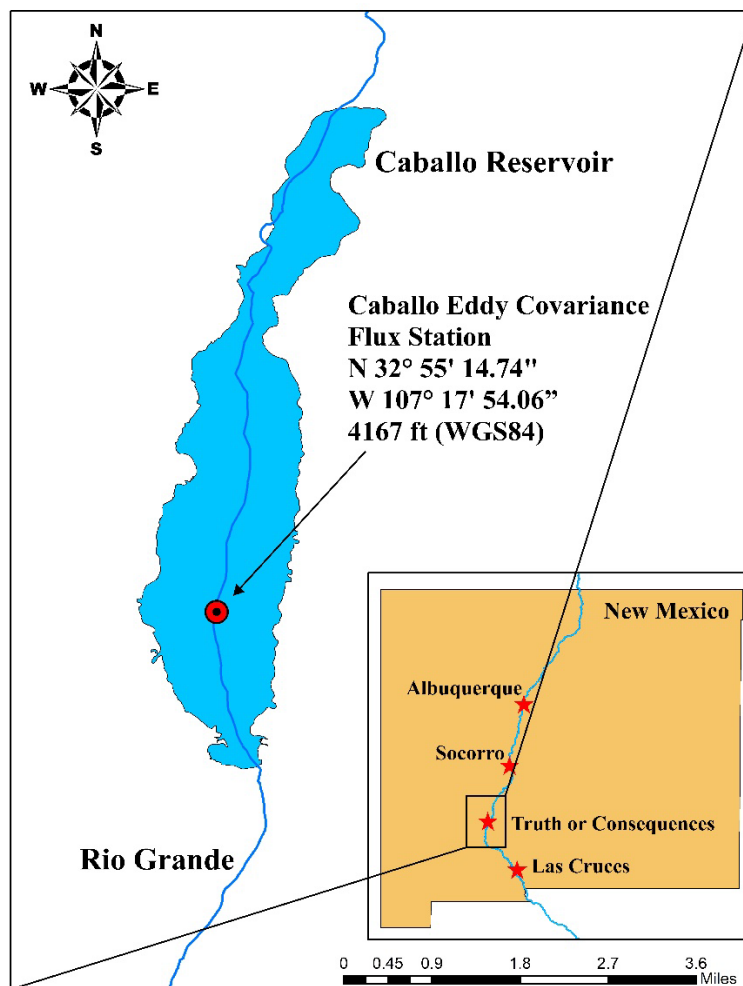

Figure 1. The geographical location of Caballo Reservoir.

4.2 Instrumentation Set up at Caballo Reservoir

4.2.1 Caballo Eddy Covariance flux station (CB-1)

Instrumentation was installed on a 50-foot triangulated tower, called Caballo Eddy Covariance flux station (CB-1) (**Figure 2**). This tower is secured with a concrete base and guy wires and it remains upright to keep the sensors level. The CB-1 location was chosen for its ideal conditions for evaporation measurements, offering a good fetch distance and capturing prevailing winds from both north-south and east-west directions. The site is easily accessible by boat and is not affected significantly by boat traffic, which minimizes artificial waves impacting the tower.

The instrumentation including sensors for bulk aerodynamic evaporation measurements are shown in the Figure 2. The meteorological sensors installed on CB-1 include a wind monitor AQ (Model 05103) for wind speed and direction, a temperature and relative humidity probe (Model HMP155A) for measuring air temperature and relative humidity. Solar radiation was measured using a pyranometer at Saltgrass climate station near the Caballo Reservoir. Data from these sensors were collected at a rate of one sample every 10 seconds and averaged every 30 minutes using a CR1000X datalogger. The datalogger, along with solar panels, battery, and enclosure were purchased from Campbell Scientific Inc., Logan, Utah.
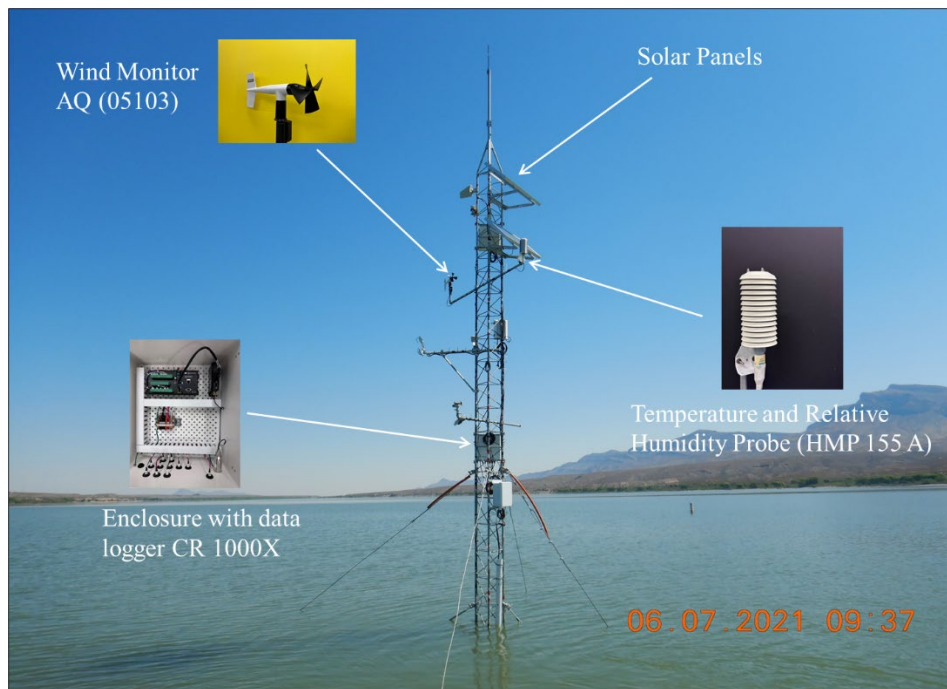

Figure 2. Instrumentation set up showing sensors of the Caballo Eddy Covariance flux station (CB-1) at Caballo Reservoir.

4.3 Adjusted Bulk-Aerodynamic Method Evaporation

Historical Caballo reservoir evaporation data measured using class A pan from 1975 to 2024 were obtained from the U.S. Bureau of Reclamation. Historical reservoir evaporation data estimated by class A pan were adjusted using a conversion coefficient determined over a three-year period to match B_AER evaporation. The conversion process involved calculating the ratio between pan evaporation and estimated bulk aerodynamic evaporation for the years 2021, 2022, and 2023. These ratios were plotted against the day of the year and fitted with a third-order polynomial function. The fitted polynomial equations derived for each year were then used to calculate daily conversion coefficients as function of the day of the year.

4.4 Long Short-Term Memory

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) designed to capture long-term dependencies in data by retaining memory of past and future information (Kisi et al., 2022). LSTM networks address the vanishing gradient problem, enabling them to learn longer-term dependencies in temporal sequences and time series data (Shi et al., 2022). Compared to traditional RNNs, the internal structure of LSTM's hidden layers is more complex (Hochreiter & Schmidhuber, 1997). LSTM networks are widely used to identify patterns in sequential data, such as text, video, speech, language, genomes, and time-series variables (Manaswi et al., 2018). LSTMs operate using special units called memory blocks, which consist of input, output, and forget gates. These memory blocks continuously update and control the flow of information (Chen et al., 2018). A typical LSTM block is composed of a forget gate, an input gate, an output gate, and a memory cell state, which runs through the entire chain to selectively add or remove information with the help of these gates (Song et al., 2020). The input gate determines whether to let new information in, the forget gate discards irrelevant information, and the output gate decides what information to output. These gates are analog gates based on the sigmoid function, operating within the range of 0 to 1 (Yadav et al., 2020). The model architecture of an LSTM is illustrated in Figure 3, and the specific procedures of the LSTM neural network at time step t are as follows:

Forget gate: $f_t = \sigma(W_f x_t + U_f h_{t-1} + b_f)$

Input gate: $i_t = \sigma(W_i x_t + U_i h_{t-1} + b_i)$

$$\tilde{C}_t = \tanh(W_c x_t + U_c h_{t-1} + b_c)$$

Cellular State: $C_t = f_t C_{t-1} + i_t \tilde{C}_t$

Output gate: $o_t = \sigma(W_o x_t + U_o h_{t-1} + b_o)$

$$h_t = o_t \tanh(C_t)$$

where W is the input weights, U is the recurrent weights, b is the bias, the subscripts *f, i, o* represents the forget gate, input gate, and output gate. The activation function of the three gates σ

is the sigmoid function which enables the valuable between 0 and 1. The activation function tanh is the hyperbolic tangent function.
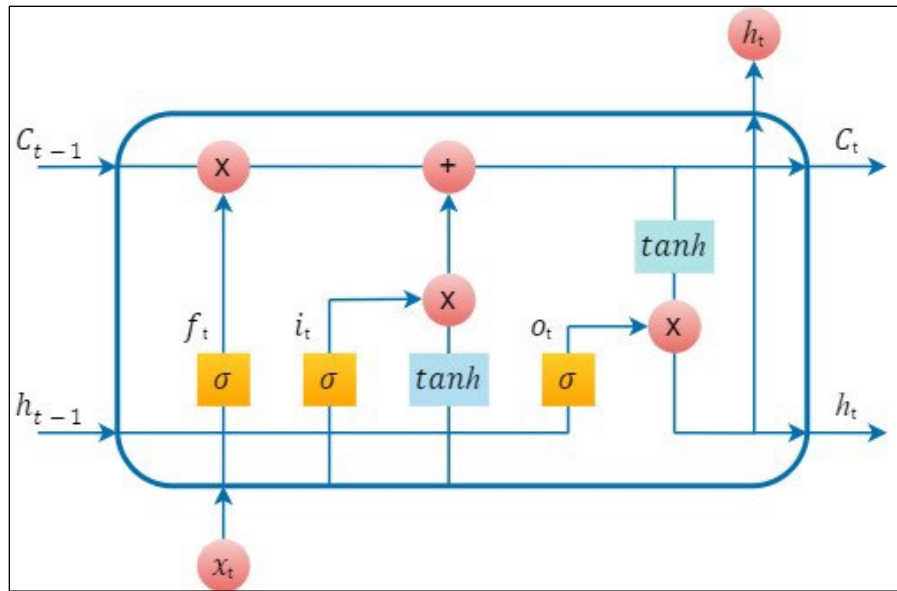

Figure 3. Typical model architecture of Long Short-Term Memory.

4.5 Model Development

The model development process for predicting evaporation using LSTM involved starting with data collection and preprocessing. Historical weather data, from January 1975 to May 2024, of daily air temperature, relative humidity, wind speed, and solar radiation were obtained from the Open-Meteo Weather API. The Meteo Weather API data from January 2021 to December 2023 were compared with measured data at CB-1 for verification. Reservoir evaporation data estimated by class A pan from January 1975 to May 2024 was obtained from the U.S. Bureau of Reclamation. The weather and the reservoir evaporation data were plotted and visually inspected to identify and remove outliers, ensuring data integrity. The percentage of missing values were very low and therefore they were ignored during the training and testing phases.

The LSTM model was constructed with a specific architecture, including a single LSTM layer with 64 neurons, utilizing a tanh activation function and a dropout rate of 0.2 to prevent overfitting. The model operated with a sequence length of 28 days. The output layer, which is the predicted evaporation, was a dense layer with one neuron and a linear activation function. The model was trained over 50 epochs with a batch size of 10 using the Adam optimizer with a learning rate of 0.01. Mean squared error (MSE) was used as the loss function during the training process. Hyperparameter tuning was performed using random search to efficiently explore the hyperparameter space and identify the optimal combination. The dataset was split into training, validation, and testing sets — 90% of the total data was used for training, of which 10% was

reserved for validation, and the remaining 10% was used for testing. Early stopping was applied to prevent overfitting by terminating the training when the validation loss stopped improving.

The model's performance was evaluated using several metrics, including root mean squared error (RMSE), mean absolute error (MAE), and R-squared (R^2), providing a comprehensive view of its accuracy and predictive power. Cross-validation was utilized to monitor and mitigate overfitting, with early stopping techniques employed to further enhance model generalization. The implementation was carried out using Python, leveraging libraries such as TensorFlow and Keras for building and training the LSTM model, with data manipulation and analysis conducted using Pandas and NumPy. The model was trained on Google Colab using the high-performance L4 GPU, known for its efficient handling of deep learning tasks.

5. Description of results; include findings, conclusions, and recommendations for further research

5.1 Climate Data

5.1.1 Air temperature

The comparison of measured air temperature at Caballo Reservoir with the FGOALS_f3_H model data from January 2021 to December 2023 reveals significant seasonal patterns (Figure 4). The FGOALS_f3_H model data were obtained from the Open-Meteo Weather API. The highest recorded temperature was 33.17°C on August 6, 2023, and the lowest was -4.86°C on February 15, 2021. Summer months (June-August) showed temperatures rise above 30°C, while winter months (December-February) often fell below 0°C, with a notable low on February 3, 2022.

The measured temperatures, shown by the solid blue line, display typical seasonal variations with peaks in summer and troughs in winter. The dashed red line, representing the FGOALS_f3_H model data, follows a similar seasonal trend, indicating strong alignment with the measured data. Minor discrepancies are observed, likely due to the model's resolution and limitations in capturing local climatic variations. This highlights the necessity of validating historical air temperature data from climate models with actual measurements to ensure accuracy and reliability.
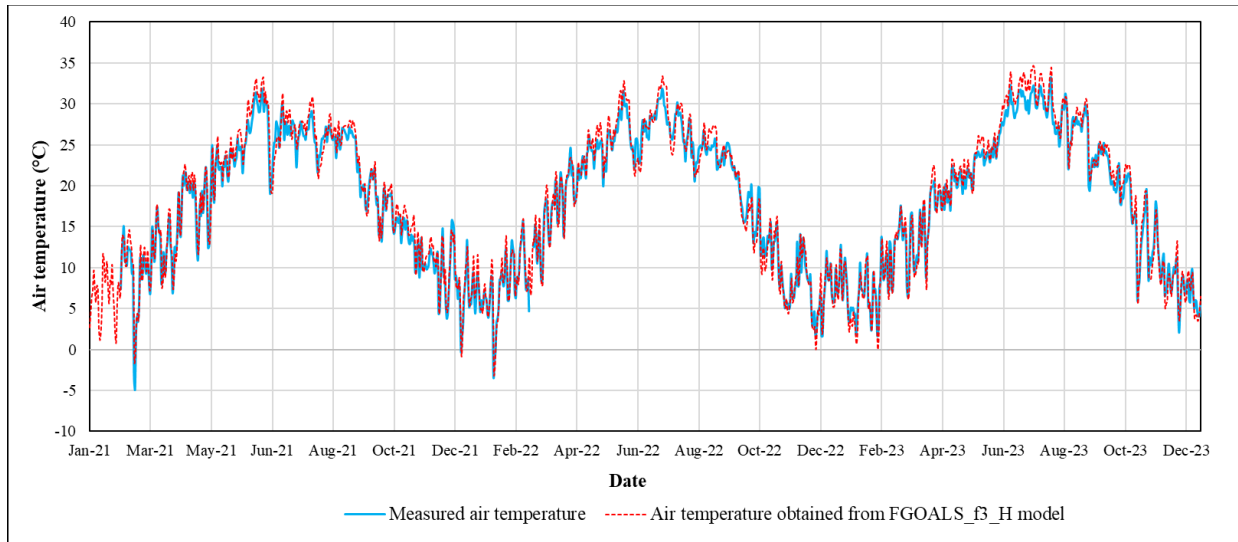
Figure 4. Comparison of air temperature measured and obtained from FGOALS_f3_H model at Caballo Reservoir from January 2021 to December 2023.

5.1.2 Relative humidity

The analysis of measured relative humidity at Caballo Reservoir compared with the FGOALS_f3_H model data from January 2021 to December 2023 shows significant seasonal trends (Figure 5). The x-axis represents the date, while the y-axis indicates relative humidity in percentage. The highest recorded relative humidity was 87.32% on December 21, 2023, and the lowest was 12.78% on April 20, 2022. Winter months (December-February) saw peaks above 80%, while summer months (June-August) experienced drops below 20%. Spring (March-May) and autumn (September-November) showed moderate humidity levels.

The measured relative humidity, illustrated by the solid green line, shows typical seasonal variations with winter peaks and summer troughs. The dashed purple line, indicating the FGOALS_f3_H model data, follows a similar pattern, suggesting a general consistency with the measured data. Discrepancies are likely due to different measurement methods: measured values are taken directly above the water level, whereas model data comes from satellite observations with a 28 km spatial resolution. This shows the need to validate historical relative humidity data from climate models with actual measurements for accurate climate assessments and projections.
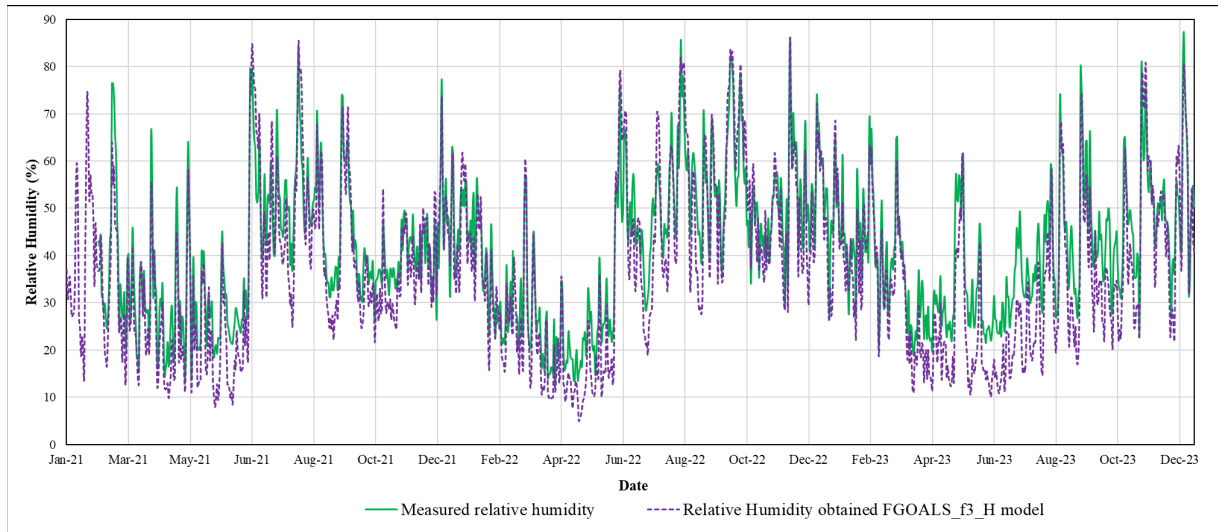
Figure 5. Comparison of relative humidity measured and obtained from FGOALS_f3_H model at Caballo Reservoir from January 2021 to December 2023.

5.1.3 Windspeed

The comparison of measured wind speed at 10 meters at Caballo Reservoir with the Open-Meteo model data from January 2021 to December 2023 highlights significant trends (Figure 6). The x-axis represents the date, while the y-axis shows wind speed at 10 meters in meters per second (m/s). The highest recorded wind speed was 11.15 m/s on February 22, 2023. Winter months (December-February) showed wind speeds often exceed 8 m/s, with a peak on February 22, 2023. Summer months (June-August) generally had lower wind speeds, frequently dropping below 2 m/s. Spring (March-May) and autumn (September-November) exhibited moderate wind speeds.

The actual measured wind speed, shown by the solid blue line, displays typical seasonal variations with winter peaks and summer troughs. The dashed red line, representing the FGOALS_f3_H model data, follows a similar pattern but tends to underestimate wind speed peaks. This discrepancy could be due to the model's resolution and limitations in capturing local climatic variations. The comparison indicates that historical wind speed data from Open-Meteo moderately reflects recent trends at Caballo Reservoir.
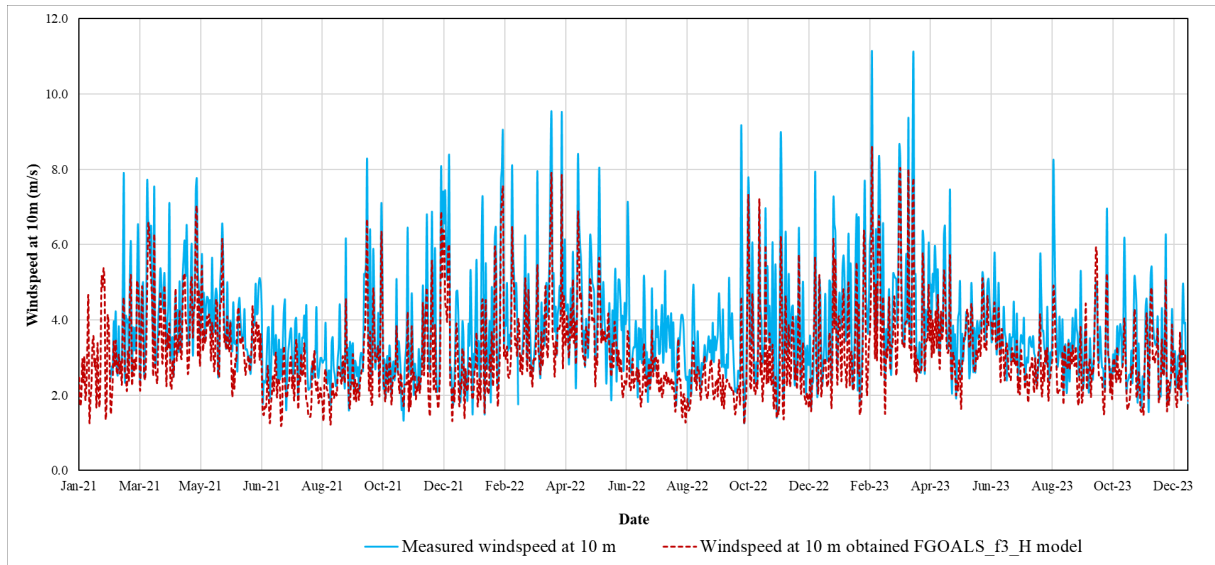
Figure 6. Comparison of windspeed measured and obtained from FGOALS_f3_H model at Caballo Reservoir from January 2021 to December 2023.

5.1.4 Solar radiation

The graph comparing measured solar radiation at Caballo Reservoir with the Open-Meteo model data from January 2021 to December 2023 shows a significant trend (Figure 7). The x-axis represents the date, while the y-axis indicates solar radiation in megajoules per square meter (MJ/m²). During this period, the highest measured solar radiation was 32.13 MJ/m² on May 24, 2021, and the lowest was 2.25 MJ/m² on December 20, 2023. High solar radiation levels, often exceeding 25 MJ/m², were recorded during summer months (June-August), with notable peaks in late May and early June each year. Conversely, winter months (December-February) typically showed lower solar radiation levels, frequently dropping below 10 MJ/m², with the lowest values recorded in late December and early January. Spring (March-May) and autumn (September-November) exhibited moderate solar radiation level.

The measured solar radiation, shown by the solid blue line, shows typical seasonal fluctuations with summer peaks and winter troughs. The dashed red line, representing the FGOALS_f3_H model data, follows a similar pattern, indicating general consistency with the measured data. Although there are notable discrepancies at a few points, with the model sometimes underestimating the peaks and overestimating the troughs, overall, the model performs well in estimating solar radiation values. This comparison highlights that historical solar radiation data from FGOALS_f3_H model data can effectively reflect recent trends at Caballo Reservoir.
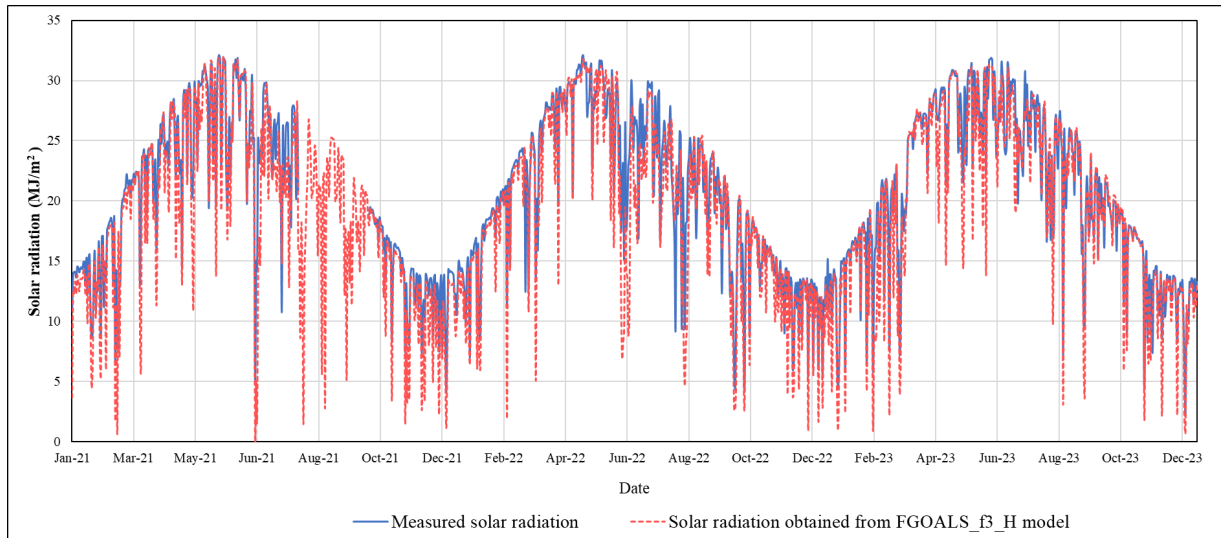
9

Figure 7. Comparison of solar radiation measured and obtained from FGOALS_f3_H model at Caballo Reservoir from January 2021 to December 2023.

5.2 Adjusted Reservoir Evaporation Estimated by Class A Pan

Figures 8 through 10 illustrate the ratio between pan evaporation and B_AER evaporation at Caballo Reservoir for the years 2021, 2022, and 2023 plotted against the day of the year (DOY) and fitted with third-order polynomial functions to determine daily conversion coefficients. Figure 8 shows fluctuations throughout the year with several peaks and troughs, indicating a relatively time series stable ratio with some variations. Figure 9 exhibits more variability in the ratio (pan evaporation/B_AER evaporation) compared to 2021, with higher peaks especially around the middle of the year, and an overall slight upward trend. Figure 10 shows a noticeable downward trend of the ratio throughout the year with significant fluctuations and higher peaks than previous 2022, suggesting greater variability and a decreasing ratio as the year progressed. The peaks and troughs are most significant in 2023, indicating greater variability in the ratio between pan evaporation and B_AER evaporation. Despite the differences in trends and variability, the calculated daily conversion coefficients using the fitted polynomial function, ranged from 0.62 to 0.83, with an average value of 0.70. The calculated coefficients remained consistent throughout the three-year period averaging to about 0.70. Consequently, an average coefficient of 0.70 was used to adjust (down scale) historical reservoir evaporation data using class A pan to B_AER evaporation for the entire historical period from 1975 to 2024. This adjusted data was considered as the actual reservoir evaporation.
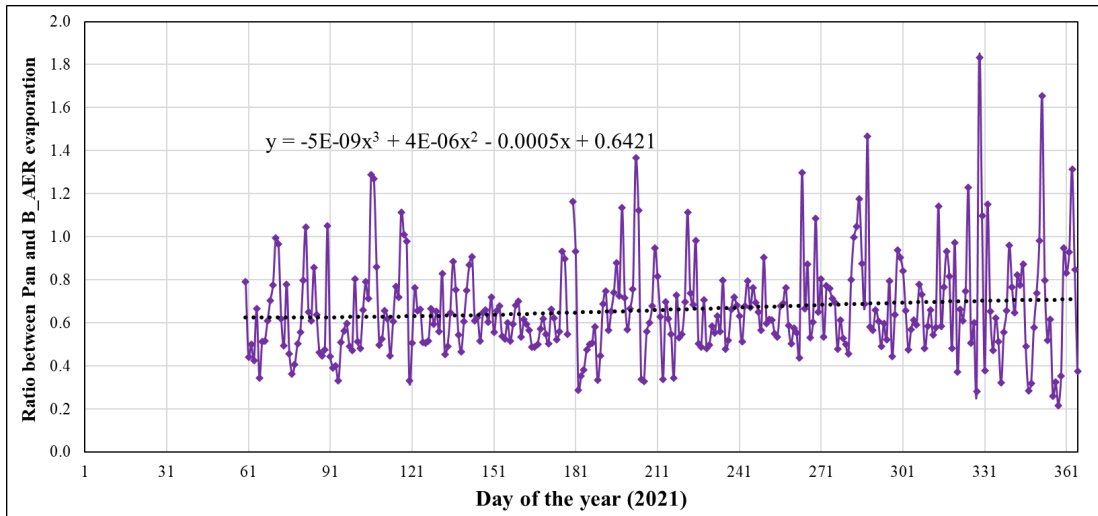
Figure 8. Ratio between pan evaporation and bulk aerodynamic (B_AER) evaporation at Caballo Reservoir for 2021, fitted with a third-order polynomial function.
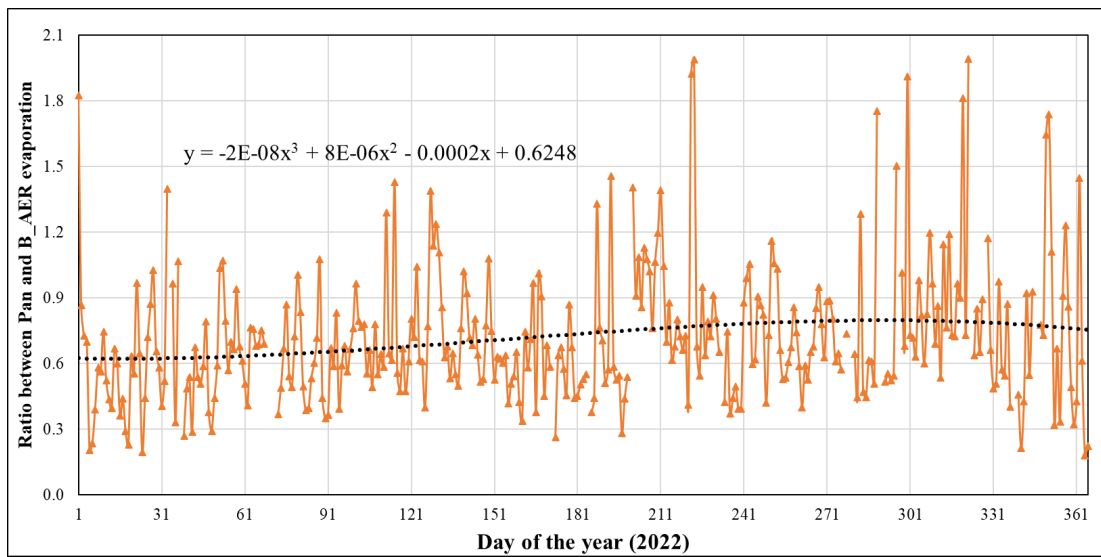

Figure 9. Ratio between pan evaporation and bulk aerodynamic (B_AER) evaporation at Caballo Reservoir for 2022, fitted with a third-order polynomial function.

Figure 10. Ratio between pan evaporation and bulk aerodynamic (B_AER) evaporation at Caballo Reservoir for 2023, fitted with a third-order polynomial function.
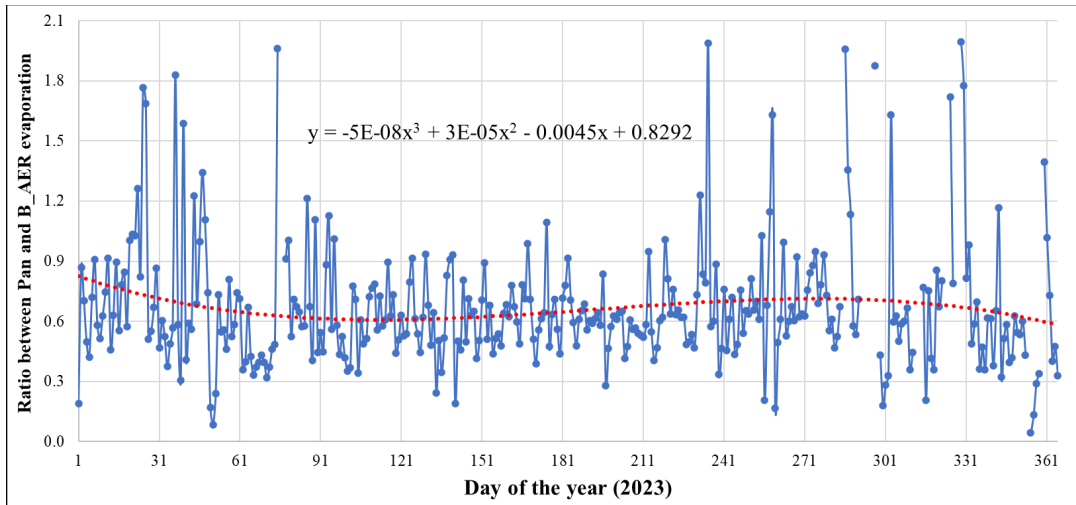
5.3 Long Term Evaporation Prediction

The analysis of the LSTM model's performance in predicting evaporation at Caballo Reservoir involves evaluating both the scatter plot (Figure 11) and the time series plot (Figure 12). These figures together provide a detailed view of the model's accuracy and its ability to capture temporal trends in evaporation. The scatter plot (Figure 11) presents predicted versus observed evaporation values. The linear regression fit is represented by the dashed black line with slope of 0.65, R^2 of 0.65 and RMSE of 1 mm. This R^2 value indicates that 65% of the variance in the observed evaporation values is explained by the model's predictions, indicating a moderate level of accuracy. The slope of the regression line is 0.65, which is less than 1, indicating that the model tends to underpredict higher evaporation values. This underprediction is more pronounced for the high observed values, highlighting a systematic bias in the model.

The time series plot of observed (solid blue line) and predicted (dashed red line) evaporation over approximately 1400 days is shown in Figure 12. The plot reveals that the LSTM model's predictions generally follow the trend of the observed evaporation, capturing the daily variations in evaporation. However, there are noticeable deviations, particularly where the predicted values do not fully match the peaks and troughs of the observed evaporation. These deviations are consistent with the underprediction tendency observed in Figure 11. The time series plot further confirms that while the model performs well in tracking overall patterns, it struggles with accurately predicting extreme values. The possible reason for the model's difficulty in predicting these peaks could be that such extreme values were not frequently encountered in the training data, leading to poor generalization in these cases. Additionally, prediction inaccuracies can occur due to unexpected external factors, such as sudden weather changes.
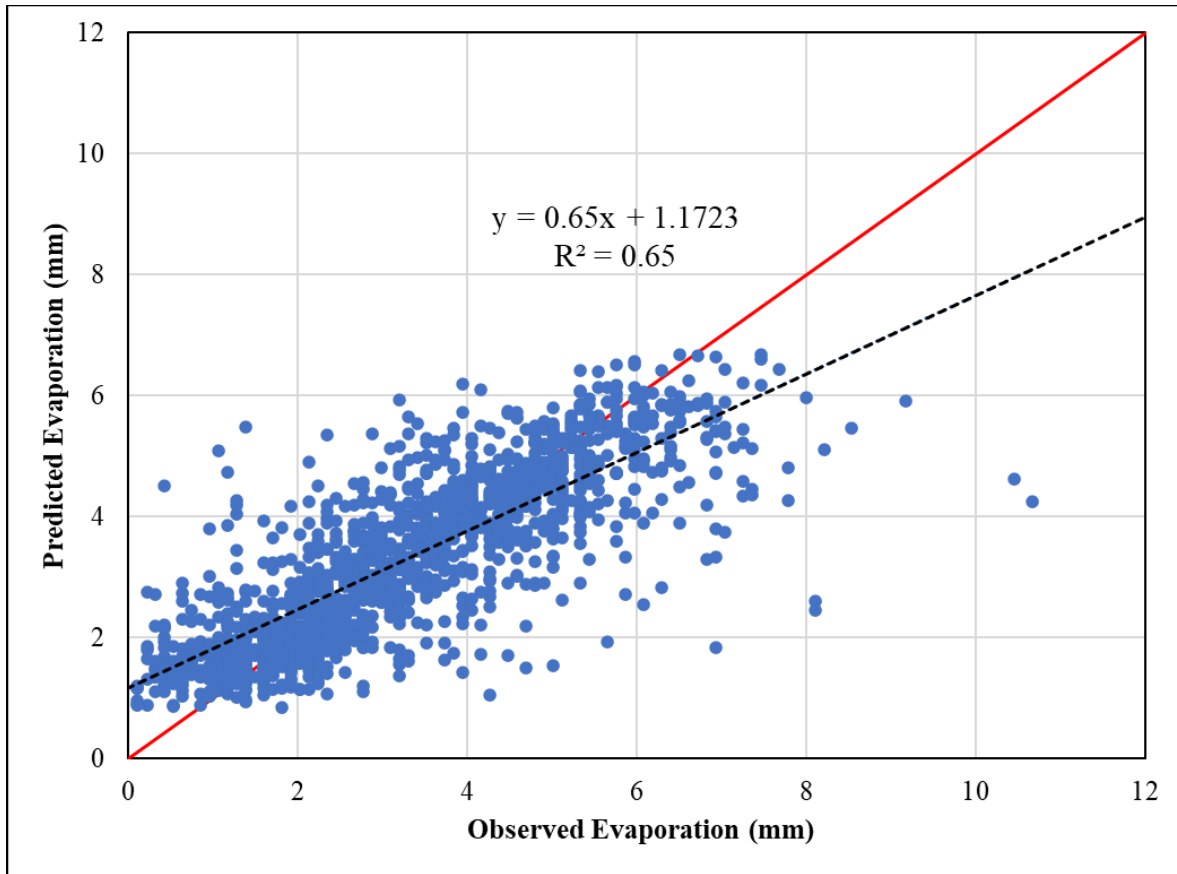
Figure 11. Scatter plot comparing observed and predicted Caballo Reservoir evaporation using the LSTM model. The plot includes the best-fit regression line (dashed black line) and 1:1 line (red).
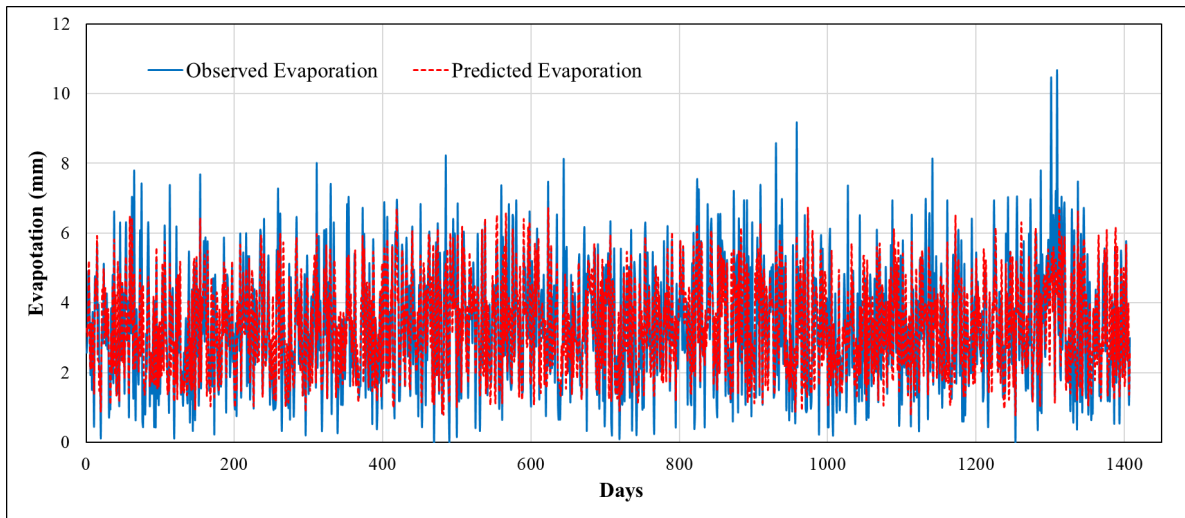

Figure 12. Time series plot showing the observed (blue solid line) and predicted (red dashed line) evaporation over 1400 days using the LSTM model.

The developed LSTM model was used to forecast the next 10 years of evaporation from 6/1/2024 to 5/31/2034 (Figure 13). The forecasted air temperature, relative humidity, wind speed, and solar radiation from the FGOALS_f3_H were used as the model input. The 10-year forecast plot shows that the model captured the seasonal cycles of evaporation, with higher evaporation occurring in the warmer months and lower values in the cooler months. The dashed red line indicates the model's predictions, which exhibit a repeating annual pattern with notable peaks and troughs consistent with expected seasonal variations.

The model's evaporation forecast highlights its usefulness of long-term evaporation projections and future water resources planning and management considering the climate change effect. However, the limitations observed in Figures 11 and 12, such as underpredicting higher evaporation values and deviations from actual data, should be considered when interpreting these forecasts. These discrepancies may be due to the model's training data not adequately representing extreme values, the inherent complexity of evaporation processes, and the model's limitation. The main limitations of this approach are its reliance on historical data, which may not capture future variability, and the risk of overfitting despite preventive techniques was implemented.
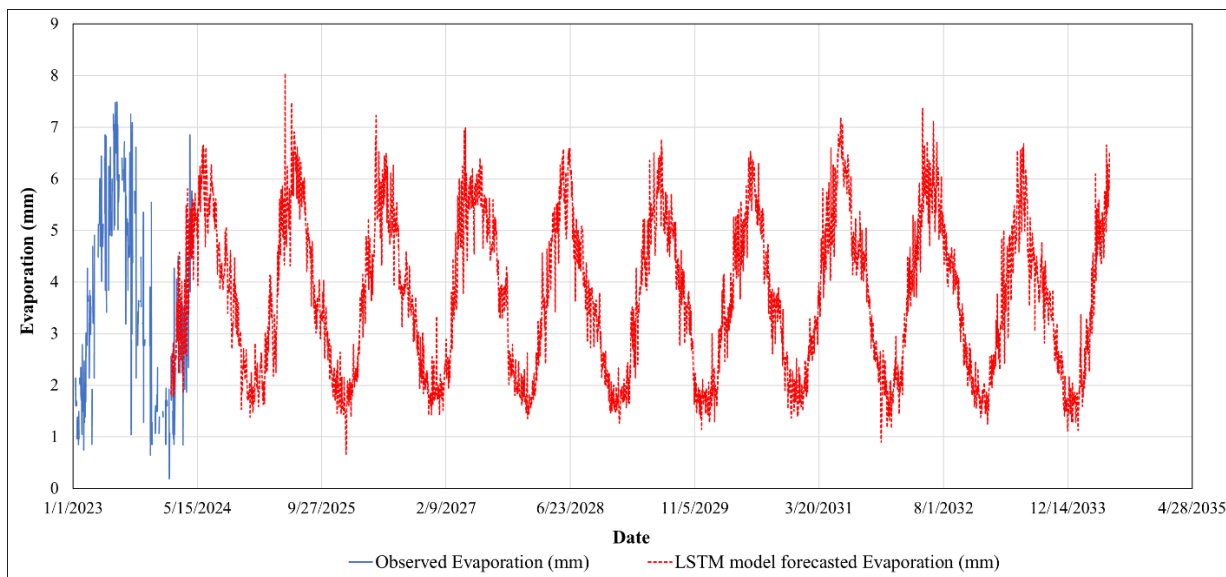

Figure 13. Forecasted evaporation for the next 10 years from 6/1/2024 to 5/31/2034 using the developed LSTM model. The plot shows observed evaporation (blue solid line) up to 5/31/2024 and model predictions (red dashed line).

6. Conclusion

The climate model FGOALS_f3_H reflected the recent weather patterns at Caballo Reservoir for air temperature, relative humidity, wind speed, and solar radiation, with minor discrepancies. The LSTM model predicted evaporation time series estimates well in general when compared to observed data. However, it did not reflect well the extreme evaporation peaks and troughs. Future evaporation forecast from 6/1/2024 to 5/31/2034 indicated the model's potential for long-term

evaporation projections and future water resources planning and management while considering the climate change effect. However, the continuous validation and adjustments are necessary to ensure reliable reservoir's evaporation predictions.

6. Provide a paragraph on who will benefit from your research results. Include any water agency that could use your results.

Farmers in the south-central part of New Mexico, El Paso, Texas, and Juarez, Mexico, rely on surface water stored in Elephant Butte and Caballo reservoirs to irrigate their crops. Therefore, the management and operation of these reservoirs are essential to the agro-economy of the region and the nation. This model will forecast evaporation rates from the Caballo reservoir. The results will help to improve the management and operation of the reservoir. Developed model with advanced techniques (AI) to predict reservoir evaporation processes will improve our understanding of the Caballo reservoir responses to a changing climate. The same methodology could be applied to other reservoirs. This model that aids policymakers, environmentalists, and researchers in water resource management and climate change mitigation strategies.

7. Describe how you have spent your grant funds. Also provide your budget balance and how you will use any remaining funds

i. Total grant funded: **$ 7,502**
ii. Expenditures:
      a. Travel expenses for the 68th Annual New Mexico Water Conference - $542.12
      b. Graduate student salary from 12/15/2023 to 01/09/2024 - $ 2,049.03
      c. graduate student's summer salary: $ 4,910.85

8. List presentations you have made related to the project.

Presented a poster at the ***68th Annual New Mexico Water Conference*** hosted by WRRI on November 8-9, 2023 at Embassy Suites by Hilton Albuquerque, New Mexico.

9. List publications or reports, if any, that you are preparing. For all publications/reports and posters resulting from this award, please attribute the funding to NM WRRI and the New Mexico State Legislature by including the account number: NMWRRI-SG-FALL2023.

None at this point

10. List any other students or faculty members who have assisted you with your project.

Anushka Perera and Juan Solis assisted with field data collection

11. Provide special recognition awards or notable achievements as a result of the research including any publicity such as newspaper articles, or similar.

None

12. Provide information on degree completion and future career plans. Funding for student grants comes from the New Mexico Legislature and legislators are interested in whether recipients of these grants go on to complete academic degrees and work in a water-related field in New Mexico or elsewhere.

Currently working on my PhD in Civil – Water Resources Engineering

References

1. Open-Meteo. (2024). *Historical Weather API*. Retrieved July 7, 2024, from https://open-meteo.com/en/docs/historical-weather-api.

2. U.S. Bureau of Reclamation. (2024). *Albuquerque Area Office*. Retrieved July 7, 2024, from https://www.usbr.gov/uc/albuq/index.html.

3. Chen, J., Zeng, G.-Q., Zhou, W., Du, W., & Lu, K.-D. (2018). Wind speed forecasting using nonlinear-learning ensemble of deep learning time series prediction and extremal optimization. *Energy conversion and management*, *165*, 681-695.

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

5. Kisi, O., Mirboluki, A., Naganna, S. R., Malik, A., Kuriqi, A., & Mehraein, M. (2022). Comparative evaluation of deep learning and machine learning in modelling pan evaporation using limited inputs. *Hydrological Sciences Journal*, *67*(9), 1309-1327.

6. Manaswi, N. K., Manaswi, N. K., & John, S. (2018). *Deep learning with applications using python*. Springer.

7. Shi, J., Jain, M., & Narasimhan, G. (2022). Time series forecasting (tsf) using various deep learning models. *arXiv preprint arXiv:2204.11115*.

8. Song, X., Liu, Y., Xue, L., Wang, J., Zhang, J., Wang, J., Jiang, L., & Cheng, Z. (2020). Time-series well performance prediction based on Long Short-Term Memory (LSTM) neural network model. *Journal of Petroleum Science and Engineering*, *186*, 106682.

9. Yadav, A., Jha, C., & Sharan, A. (2020). Optimizing LSTM for time series prediction in Indian stock market. *Procedia Computer Science*, *167*, 2091-2100.