Article

Enhancing Monthly Streamflow Prediction Using Meteorological Factors and Machine Learning Models in the Upper Colorado River Basin

Saichand Thota [1], Ayman Nassar [2,3], Soukaina Filali Boubrahimi [1,*], Shah Muhammad Hamdi [1] and Pouya Hosseinzadeh [1]

[1] Department of Computer Science, Utah State University, Logan, UT 84322, USA; s.thota@usu.edu (S.T.); s.hamdi@usu.edu (S.M.H.); pouya.hosseinzadeh@usu.edu (P.H.)
[2] Utah Water Research Laboratory, Department of Civil and Environmental Engineering, Utah State University, Logan, UT 84322, USA; ayman.nassar@usu.edu
[3] Department of Civil and Environmental Engineering, University of Utah, Salt Lake City, UT 84112, USA
[*] Correspondence: soukaina.boubrahimi@usu.edu

Abstract: Streamflow prediction is crucial for planning future developments and safety measures along river basins, especially in the face of changing climate patterns. In this study, we utilized monthly streamflow data from the United States Bureau of Reclamation and meteorological data (snow water equivalent, temperature, and precipitation) from the various weather monitoring stations of the Snow Telemetry Network within the Upper Colorado River Basin to forecast monthly streamflow at Lees Ferry, a specific location along the Colorado River in the basin. Four machine learning models—Random Forest Regression, Long short-term memory, Gated Recurrent Unit, and Seasonal AutoRegresive Integrated Moving Average—were trained using 30 years of monthly data (1991–2020), split into 80% for training (1991–2014) and 20% for testing (2015–2020). Initially, only historical streamflow data were used for predictions, followed by including meteorological factors to assess their impact on streamflow. Subsequently, sequence analysis was conducted to explore various input-output sequence window combinations. We then evaluated the influence of each factor on streamflow by testing all possible combinations to identify the optimal feature combination for prediction. Our results indicate that the Random Forest Regression model consistently outperformed others, especially after integrating all meteorological factors with historical streamflow data. The best performance was achieved with a 24-month look-back period to predict 12 months of streamflow, yielding a Root Mean Square Error of 2.25 and R-squared (R^2) of 0.80. Finally, to assess model generalizability, we tested the best model at other locations—Greenwood Springs (Colorado River), Maybell (Yampa River), and Archuleta (San Juan) in the basin.

Keywords: streamflow prediction; machine learning; time series regression; upper Colorado river basin; RFR; snow water equivalent

1. Introduction

Rivers are complex, multiattribute, and multifunctional systems [1]. Their circulation, development, and uses are comprehensively influenced by human activities and climatic changes [2]. The water management community has long been interested in improving streamflow forecasts, especially for long-term planning. They also seek to refine the understanding of streamflow forecast accuracy and how to interpret and utilize these forecasts effectively in reservoir operations. Ultimately, their goal is to develop operating policies that optimize the use of hydrological forecasts [3]. Accurate forecasts substantially improved reservoir operations in reservoirs that operate to meet a target water elevation [4]. This prediction plays a crucial role in various water resource applications such as optimizing hydropower generation, managing reservoirs efficiently, allocating water resources

effectively, protecting the environment [5], and flood warning [6]. Hydrologic time series forecasting holds significant importance in operational hydrology and has garnered considerable attention from researchers over the last few decades.

Numerous models have been proposed to enhance hydrologic forecasting, reflecting the ongoing efforts to improve the accuracy and reliability of predictions in this field [7–9]. Streamflow prediction can be conducted in four ways: 1. Physical-based models: they involve simulating the movement of water through a watershed using principles of physics, hydrology, and fluid mechanics; 2. Conceptual and Semi-distributed models: they involve employing simplified representations of hydrological processes to forecast river discharge; 3. Data-driven models: they involve utilizing historical data on factors like precipitation, temperature, and land cover to forecast future streamflow levels; 4. Metric models, such as IHACRES (Identification of unit Hydrographs And Component flows from Rainfall, Evaporation, and Streamflow data) [10], do not rely on specific hydrological features or processes but instead are built upon unit hydrograph theory [5]. Physical models require extensive and precise data for parameter calibration, which can be challenging to obtain. Due to these limitations, acquiring sufficient accurate data becomes difficult, resulting in suboptimal model performance and increased uncertainty [11]. A conceptual model focusing on the components may yield better results than a physical model [12]. However, including too many component processes in the model can lead to overparameterization, increasing the risk of introducing unnecessary complexity [13] and their effectiveness may vary depending on the values of parameters and assumptions used. A primary drawback of metric models lies in their dependency on simplifications and assumptions regarding watershed behavior, which may not comprehensively represent the complexities of hydrological processes. Furthermore, these models are also sensitive to parameterization and struggle to accommodate non-stationary data, such as changing climate conditions. Machine learning (ML) models, such as the gated recurrent unit (GRU) [14], and long short-term memory network (LSTM) [15], are well-suited for handling highly volatile and nonlinear data, such as monthly runoff [16]. ML models typically involve mapping multiple input features to output targets [17]. These data-driven techniques typically demand comparable data, as in the aforementioned models, but necessitate significantly less development time. They are particularly beneficial for real-time applications and demonstrate proficiency in accurately predicting streamflow [18,19].

A significant hurdle in streamflow forecasting is determining an effective approach for predicting streamflow over timeframes, which is crucial for aiding water resource managers and decision-makers in understanding the system and planning for future management strategies. Forecasting streamflow at various intervals, such as hourly, daily, monthly, and yearly, holds great importance in optimizing water usage across different applications. Nagar et al. [20] emphasize the significance of hourly streamflow predictions over monthly data, highlighting its importance for evacuation and flood planning. In contrast, Hosseinzadeh et al. [21] prioritize monthly predictions for their utility in river construction and development projects. Determining the optimal timeframe for streamflow prediction is crucial and largely depends on the available data and the specific objectives of the forecasting task. Typically, the timeframe for streamflow prediction should balance between the need for timely information and the desire for accurate forecasts.

In recent years, ML models have gained popularity in streamflow forecasting due to their ability to capture complex relationships within hydrological data [18,22]. Furthermore, monthly streamflow prediction tends to yield better results in wet regions compared to dry regions [23]. Different ML methods have been used for predicting streamflow, one of which is Artificial Neural Networks (ANNs) [24,25]. These models are flexible and can understand complex connections between input factors and streamflow, making them successful for short-term predictions [26]. Support Vector Machines (SVMs) [27] are effective for handling high-dimensional data and can provide accurate streamflow predictions, especially with small datasets [28]. Random Forests Regression (RFR), an ensemble robust learning method, can capture complex interactions among predictors, making it suitable for streamflow

forecasting tasks [29]. Recurrent Neural Networks (RNNs), RNNs, including LSTM and GRU networks, excel in capturing temporal dependencies in sequential data, making them well-suited for streamflow time series prediction [30]. SARIMA (Seasonal AutorRgressive Integrated Moving Average) model [31] is a time series forecasting method that extends the ARIMA (AutoRegressive Integrated Moving Average) model [31], by including seasonal patterns in the data. SARIMA models are useful for capturing and predicting complex temporal patterns in data with seasonality, making them commonly applied in various fields, including economics, finance, and hydrology, for forecasting time series data with recurring seasonal patterns [32,33].

For time series forecasting, data need to be reshaped to utilize standard linear and nonlinear ML algorithms. This reshaping process commonly employs the sliding window method, where a sequence of time series data is transformed [34–38]. Here, the input variables consist of values from previous time steps, while the output variable corresponds to the value at the next time step. Thus, time series forecasting is reformulated as an ML problem, leveraging the value at the previous time step to predict the value at the next time step [39–41]. Overall, ML models offer promising opportunities for improving streamflow forecasting accuracy and can complement traditional hydrological models in capturing the complex dynamics of hydrological systems. However, careful model selection, feature engineering, and validation are crucial to ensure robust and reliable predictions.

Streamflow prediction has evolved significantly with the integration of ML models such as RFR, LSTM, and GRU networks, alongside traditional methods like SARIMA. These advancements are crucial for hydrologists to select the most effective models for forecasting streamflow. Research by Gao et al. [18] highlighted the superiority of LSTM and GRU models over ANN for short-term (hourly) runoff predictions at the Yutan station control catchment in Fujian Province, Southeast China. Hosseinzadeh et al. [21] found in their study of the Upper Colorado River Basin that the performance order of ML models for monthly streamflow prediction was RFR, SARIMA, and LSTM. These models utilize historical streamflow data to forecast future streamflow values, which helps in understanding the streamflow variability under different climatic conditions. The time series analysis method, SARIMA, while widely used, struggles to capture nonlinear relationships among hydrological variables and requires stable hydrological series, limiting its practical applicability [42]. Therefore, there is a growing interest in comparing the performance of SARIMA with other models to understand their differences in capturing seasonality. In related research, multivariate analysis, such as snow flow prediction using hybrid models trained with meteorological data, has proven to yield better results compared to models that do not include such data [43]. Furthermore, studies have demonstrated improvements in streamflow predictions by incorporating atmospheric circulations and considering Pacific sea surface temperature in the Upper Colorado River Basin [44]. The Colorado River catchment is currently experiencing significant challenges, notably a shortage of water, which has become a contentious issue in recent years. Dry conditions over the past two decades, compounded by global climate change, are impacting this huge catchment, which spans several states in the United States. Studies indicate that the Colorado River basin may experience a reduction in runoff of around 19% by the middle of the 21st century [45–47]. Moreover, prolonged rainfall events in certain seasons can lead to flooding in the area [48]. Given these circumstances, timely decision-making and proactive measures have become crucial. Streamflow prediction in this region would be instrumental in helping water resource managers and basin stakeholders mitigate the risks of disasters and effectively manage water resources.

These findings highlight the importance of integrating various environmental factors and leveraging advanced modeling techniques to enhance streamflow prediction accuracy. Therefore, in our study, we focused on predicting monthly streamflow by considering both flow and meteorological factors using ML models such as RFR, LSTM, GRU, and SARIMA. Among various factors such as temperature, precipitation, snow water equivalent, snow depth, snow density, and soil moisture level, we found that temperature, precipitation,
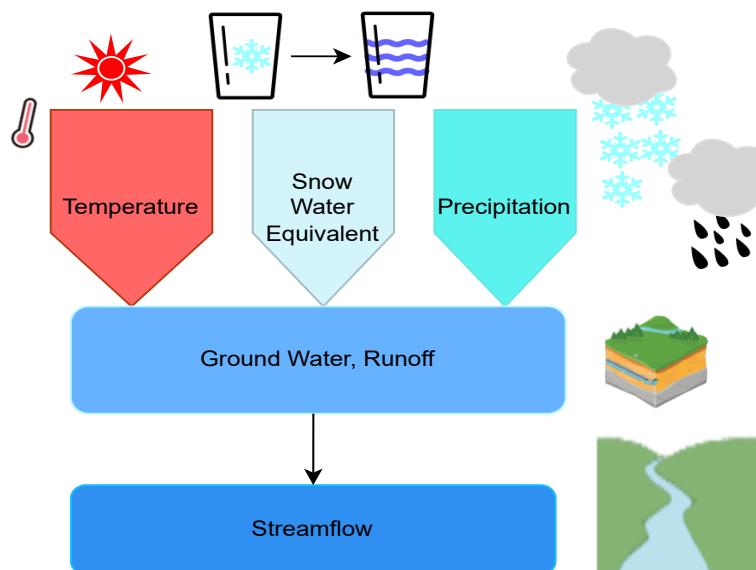
Figure 1. Visual Representation of Analyzed Factors (Temperature, Snow Water Equivalent, Precipitation, and Streamflow) in Streamflow Prediction.

2. Study Site

The Colorado River is the most overallocated river in the world [49]. It provides water for nearly 40 million people in the southwestern United States and northern Mexico. It is projected to be greater than supply by approximately 4×109 m^3 in the year 2060 [50]. High water demand, decades of national and international treaties, and over 40 major dams make the Colorado River Basin (CRB) perhaps the most regulated watershed on Earth [46]. Historically, management of water resources in the Colorado River Basin focused largely on surface water [51]. It is a lifeline for the population and agricultural economy of parts of seven U.S. states (Wyoming (WY), Utah (UT), Colorado (CO), New Mexico (NM), Nevada (NV), Arizona (AZ), and California (CA)) and the Mexican states of Sonora and Baja California [52]. The river basin is divided into Upper and Lower Basin, with Lees Ferry as the dividing point. The Upper Basin serves the states of WY, CO, UT, and NM. The Lower Basin serves the states of NV, AZ, and CA within the United States as well as Mexico. High-elevation snowpack in the Rocky Mountains contributes about 70% of the annual runoff, and the seasonal runoff pattern throughout most of the basin is heavily dominated by winter snow accumulation and spring melt [49]. Roughly 90% of the river's flow is derived from snowmelt from precipitation in three upper basin states, Colorado, Utah, and Wyoming. However, most of the demand and use of the flows are in the lower basin states, Arizona, California, and Nevada [53].

In this study, we focused only on the Upper Colorado River Basin (UCRB), Figure 2. UCRB is a snow-melt-dominated hydrologic system that covers about 280,000 km^2. It extends from headwaters in the Rockies in Colorado and Wyoming to Lee's Ferry in Northern Arizona with elevation ranging between 3300 m and 900 m. During the winter season, from October to the end of April, the snow cover area (SCA) for the UCRB ranges from 50,000 km^2 to 280,000 km^2 which plays a crucial role in energy [54] and hydrological [55]

cycles. The primary stream in the UCRB is the Upper Colorado River, with major tributaries including Williams Fork, Blue River, Muddy Creek, Eagle River, Roaring Fork River, Rifle Creek, Gunnison River, Plateau Creek, and Fraser River [21].

Despite little change in precipitation in the Upper Colorado River Basin (UCRB) between 1896 and 2019, temperatures have risen [56] and water supplies in the basin have suffered [57]. The warmer air temperatures are connected to decreases in streamflow and shifts in snowmelt-runoff timing to earlier in the Spring, thereby depleting streamflow during the Summer season at the peak of water demands [58–62]. Climate variables such as regional precipitation (rainfall and snowfall) and snowpack have large impacts on streamflow. These variables have been applied to short-lead seasonal predictions of streamflow and water supply for the Colorado River, including those from the Natural Resources Conservation Service and Colorado Basin River Forecast Center. Although snow water equivalent in April has the dominant influence on peak flow of the UCRB in April–July, precipitation in spring can significantly influence snow melting and runoff of the UCRB, and so influence the year-to-year variation of the UCRB streamflow [63–65].
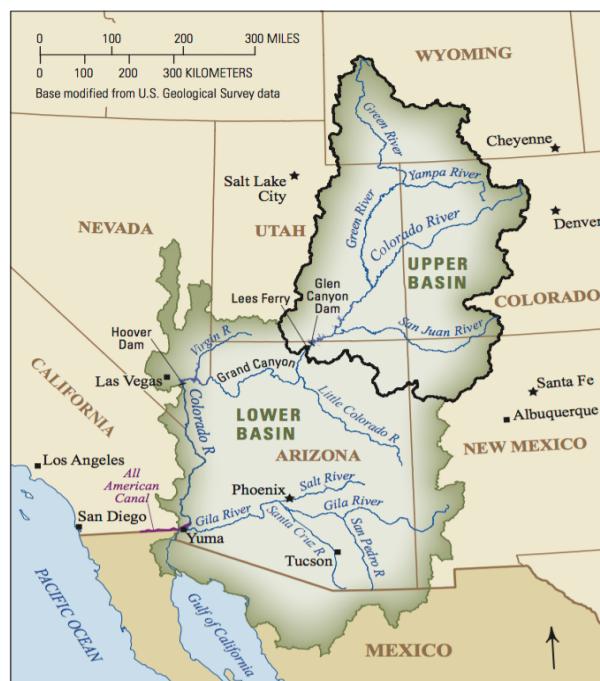
Figure 2. Colorado River Basin by United States Geological Survey (USGS) (https://www.usgs.gov/media/images/colorado-river-basin-map (accessed on 10 February 2024)).

3. Data

Our research hinged on two key datasets: historical streamflow records for the Colorado River and environmental data derived from the network of Snow Telemetry (SNOTEL) stations. Our investigation was particularly confined to the Upper Colorado River Basin (UCRB), and our primary objective was to analyze the SNOTEL data comprehensively and employ it in the prediction of streamflow within this specific region.

The streamflow dataset was meticulously compiled from the United States Bureau of Reclamation (USBR) website, accessible at (https://www.usbr.gov/lc/region/g4000/NaturalFlow/current.html, accessed on 19 September 2023). This dataset spanned a considerable time frame, encompassing 115 years, starting from 1905 and concluding in 2020. Monthly measurements of river discharge were recorded at a total of 29 distinct monitoring locations. These stations included well-known sites such as the Colorado River At Lees Ferry, AZ, the Yampa River Near Maybell, CO, and the Colorado River Below Parker Dam, AZ-CA. To better focus our analysis on the UCRB, we accurately delineated the geographical boundaries of this region, using geospatial data provided by the ScienceBAse-

Catalog of United States Geological Survey (USGS), website—(https://www.sciencebase.gov/catalog/item/imap/4f4e4a38e4b07f02db61cebb, accessed on 19 September 2023). This effort enabled us to identify a subset of 20 monitoring stations called USGS gauge locations, that fell within the UCRB's domain. For our analysis, we standardized the streamflow values by converting them into millimeters per month (mm/month) through a normalization process based on the UCRB's area. This transformed representation was used consistently in our research to facilitate comparative analysis and modeling.

In parallel, we turned our attention to the SNOTEL network, a system of automated monitoring stations positioned throughout the United States. These stations are primarily operated by government agencies, with the United States Department of Agriculture's Natural Resources Conservation Service (USDA NRCS) and the National Oceanic and Atmospheric Administration (NOAA) playing important roles. The SNOTEL network is instrumental in collecting data related to snowpack, precipitation, temperature, and various other meteorological and hydrological parameters. The SNOTEL network's data collection strategy collects snow-related and real-time data, as these sensors had been deployed allowing us access to data spanning 30 years, leading up to December 2020. As of December 2020, our analysis revealed the existence of 138 SNOTEL sensors within the boundaries of the UCRB, according to data from the USDA NRCS, website accessible at [60]. The USDA NRCS website also provides a report generator, accessible at https://wcc.sc.egov.usda.gov/reportGenerator/, accessed on 19 September 2023), allowing users to retrieve various data concerning SNOTEL sensors. These sensors provided a diverse range of data types, including measurements of snow water equivalent, snow depth, precipitation, temperature, snow density, and soil moisture. Snow water equivalent is how much depth water would cover the ground if the snow cover were to melt and become a liquid. The data were available at different temporal resolutions, ranging from hourly to daily, monthly, and even half-yearly intervals. To ensure that the SNOTEL data were compatible with the streamflow data, we processed and aggregated the daily SNOTEL records, which often amounted to more than 46,000 data values, into a consistent monthly format. This alignment allowed for more effective comparative analysis and facilitated the integration of these datasets. Figure 3 provides a clear visual representation of the distribution of SNOTEL sensors within the UCRB and their relation to the USGS monitoring gauges. In this paper, the terms 'streamflow' and 'flow' are used interchangeably and refer to the same hydrological parameter.
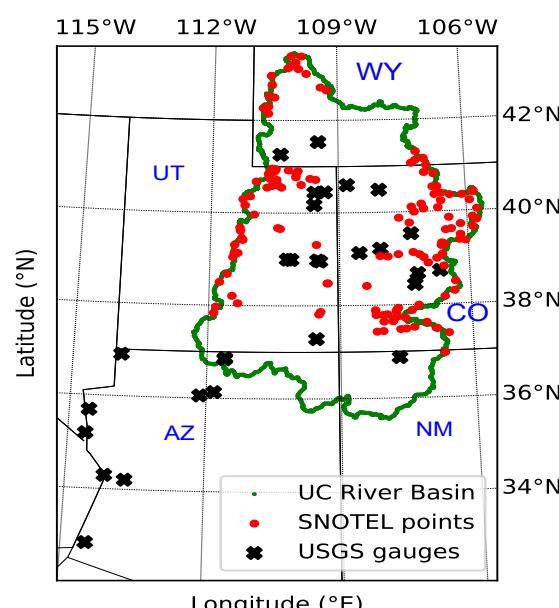

Figure 3. Spatial Distribution of SNOTEL Weather Stations and USGS Gauges within the Upper Colorado River Basin.

4. Data Analysis

In this section, we present an overview of the data analysis conducted for streamflow forecasting. This encompasses preprocessing the data acquired from both SNOTEL and USBR sources, selecting relevant features, representing data over the specified time, and dividing the dataset for training and testing purposes.

We extracted data from SNOTEL sensors, including Snow Water Equivalent, Snow Depth, Precipitation Accumulation, Temperature (observed, max, min, and average), Snow Density, and Station Name. Upon analysis, it was observed that there are numerous null values in attributes like Snow Depth (over 16,000) and Snow Density (almost 25,000), despite the data being collected from sensors since 1991. However, there is a lack of continuous data for Snow Depth and Density until 2008. Consequently, these attributes, 54% missing values for snow depth and 74% missing values for snow density were omitted from our analysis, focusing on the available sensor data rather than prediction. For temperature-related data, we prioritized the average temperature as the primary feature, discarding the observed, minimum, and maximum attributes. Ultimately, our analysis concentrated on Snow Water Equivalent (SWE), Precipitation Accumulation (Prcp_Acc), and Average Temperature (Temp_Avg) as the main data derived from the SNOTEL sensors. Since the streamflow data are available on a monthly basis, we aggregated the SNOTEL data on a monthly scale while excluding stations. After monthly data aggregation, no missing values were found for temperature, precipitation, and snow water equivalent. However, snow depth was missing 24% of its values, and snow density was missing 47.5%. Therefore, it is reasonable to exclude these two attributes from the analysis. Consequently, we obtained a data frame with 360 rows, representing 30 years of data. Time series plots of these attributes along with the moving average are plotted over the years, as illustrated in Figure 4 while the monthly streamflow data obtained from the USBR is depicted below in Figure 5. The time series plot of average temperature exhibits a slight positive trend, indicative of the global warming pattern observed in the UCRB region. To facilitate model training and evaluation, we split the 30-year dataset into 80% for training purposes and 20% for testing. The performance of the models will be assessed by comparing their predicted values with the actual streamflow values obtained from the USBR, using evaluation metrics.
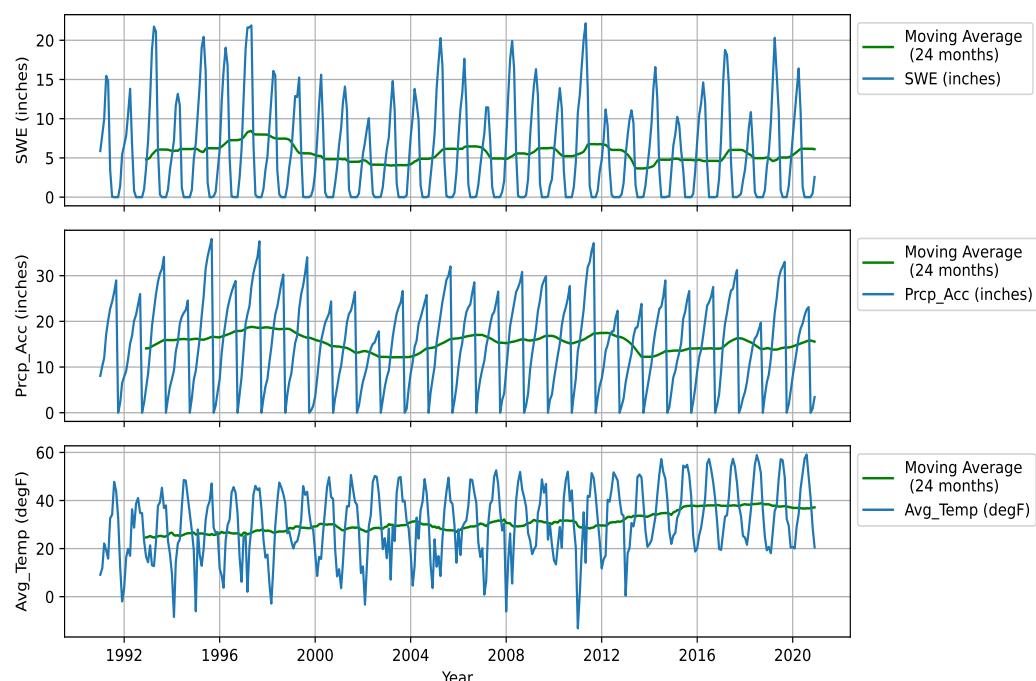

Figure 4. Time Series Plots of SNOTEL Data - Snow Water Equivalent (SWE) in inches, Precipitation Accumulation (Prcp_Acc) in inches, and Average Temperature (Avg_Temp) in degrees Fahrenheit.
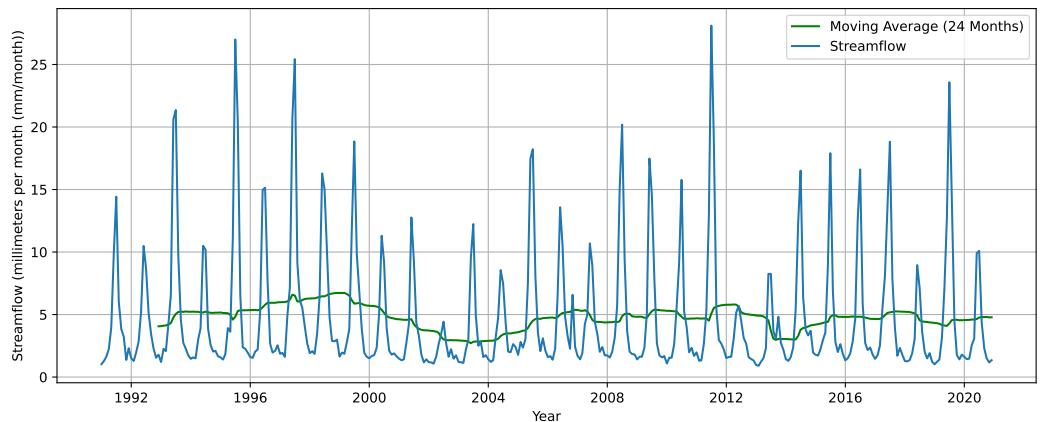
Figure 5. Time Series Plot of Streamflow Data from the US Bureau of Reclamation.

In our analysis, we incorporated the streamflow data from past months to predict future months. Additionally, we explored the correlation between these attributes in Figures 6 and 7 to illustrate the variations in streamflow corresponding to each attribute. We used Pearson correlation in our analysis, as it is widely used in scientific research due to its effectiveness in assessing linear relationships between variables. It measures the strength and direction of such associations, making it ideal for datasets where linearity is assumed. The obtained results align with expectations, revealing a high correlation between the predicted value, streamflow, and the average temperature. Since we are analyzing data from a large area, we cannot completely rely on the correlation between snow water equivalent and temperature. Snow water equivalent is influenced by temperature and precipitation at higher altitudes, but at lower levels, it is heavily affected by snow depth [67] At higher altitudes, the high temperature leads to a decrease in snow water equivalent and an increase in precipitation. In aggregate, these factors contribute to an overall increase
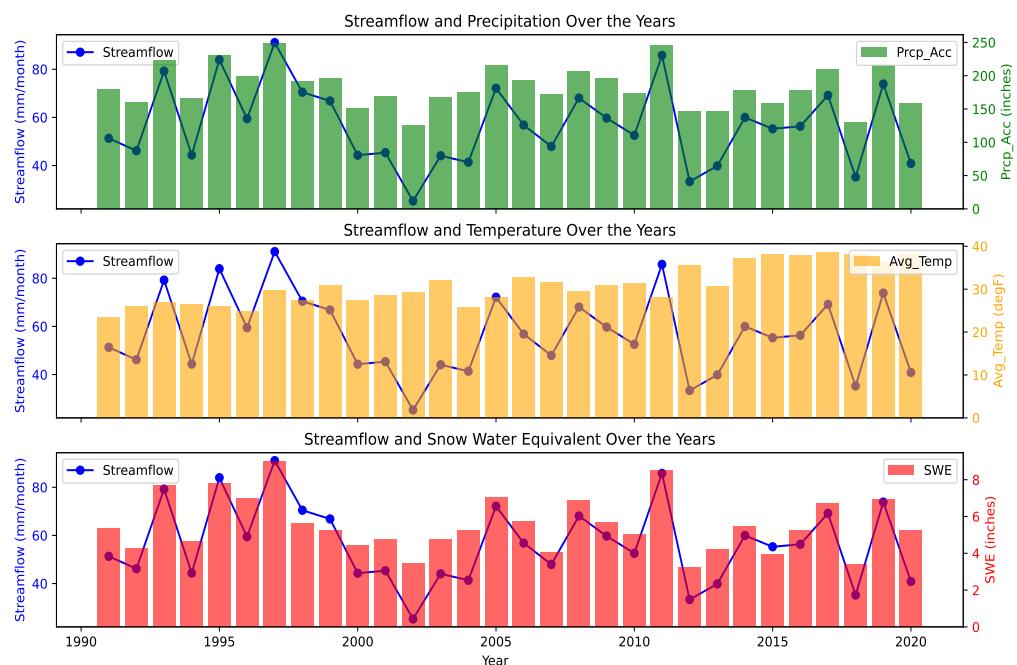
Figure 6. Temporal Variation of Streamflow in Relation to Meteorological Variables.

Notably, the removal of the streamflow value during feature selection yielded variations in the predicting percentage, as elucidated by comparing the results in Section 6, Experimental Results.
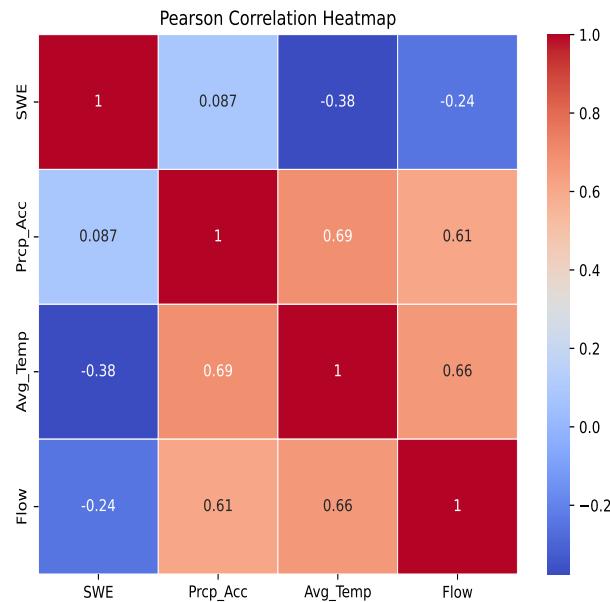
**Figure 7.** Correlation Analysis between Streamflow and Meteorological Variables.

5. Methodology

This section provides an overview of the methodologies employed for streamflow forecasting. We delve into the mathematical expressions defining the models and discuss the evaluation metrics. The section is structured into three subsections for clarity: (1) univariate time-series prediction model, (2) multivariate time-series prediction model, and (3) evaluation metrics.

5.1. Univariate Time-Series Prediction Model

In the univariate time series prediction model, both the input and output features are the same, focusing on the streamflow without introducing additional variables. Specifically, the Lees Ferry streamflow has been selected for prediction, as it serves as the boundary between the Upper and Lower Colorado River Basin (CRB). Machine learning models utilized for this prediction task include RFR, LSTM, SARIMA, and GRU.

The entire dataset has been partitioned into sets of look-back (past) and look-ahead (future) sequences for training and testing purposes. Given the limited data availability, encompassing only 30 years, with 6 years allocated for testing, various sequence lengths have been considered, including 12, 24, 36, 48, and 60 months, summing up to a maximum of 72 months (6 years) for any given combination of input and output sequence. Optimal results were achieved with a configuration using 12 months of look-back and 12 months of look-ahead data for uni-variant data. Contrary to common normalization practices such as Z-Normalization or Min-Max Scaling, Equations (1) and (2), respectively, it was observed that applying such transformations did not yield significant improvements when compared to using the actual, unaltered data.

$$Z_{norm} = \frac{x - \mu}{\sigma} \tag{1}$$

where x is the observed value, μ and σ are the mean and standard deviation of the data.

$$x_{scaled} = \frac{x - x_{min}}{x_{max} - x_{min}} \tag{2}$$

where x is the observed value, and x_{min} and x_{max} are the minimum and maximum values for that particular attribute.

Consequently, the analysis proceeded without normalization, employing the data after preprocessing for RFR, LSTM, and GRU models. SARIMA models require the entire dataset to estimate their parameters. Therefore, in SARIMA modeling, the entire dataset was utilized for training the model instead of using sequences. To ensure fairness in result comparisons and observations, the same dataset was employed for both training and testing across all combinations of models and sequence lengths.

5.1.1. RFR

A Random Forest is a meta-estimator that involves fitting multiple decision tree regressors on different subsets of the dataset. The algorithm utilizes averaging to enhance predictive accuracy and mitigate the risk of overfitting. Figure 8 illustrates a simplified structure of RFR where each tree in the ensemble has a depth of 2. In general, RFR can be defined as the following equation:

$$RFR(x) = \frac{1}{N}(T_1(x) + T_2(x) + \ldots + T_N(x)),$$ (3)

where N is the number of decision trees, Ti(x), is the prediction made by the tree *i* on input

mproved
er values
f set to 1,
igated to
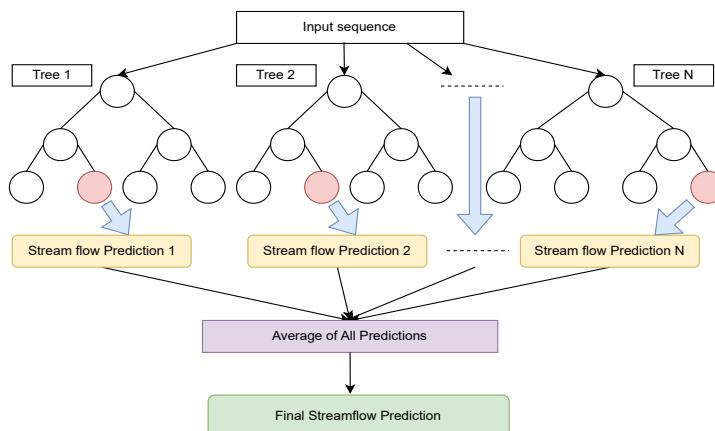
Figure 8. A Simplified Structure of RFR.

5.1.2. LSTM

Long Short-Term Memory (LSTM) is a type of recurrent neural network (RNN) architecture designed to address the vanishing gradient problem in traditional RNNs, using the below architecture, Figure 9. LSTMs are particularly well-suited for processing and predicting sequences of data due to their ability to capture and remember long-term dependencies. They are widely used in natural language processing, speech recognition, time-series forecasting, and other tasks where capturing long-term dependencies is crucial. Activation functions, such as sigmoid, tanh, Rectified Linear Unit (ReLU), and softmax, are mathematical operations utilized in neural networks to introduce non-linearity, facilitating the learning of complex patterns in data. Specifically within LSTM cells, sigmoid layers regulate information flow by selectively gating input, output, and forget signals. These layers utilize activation functions to produce outputs between 0 and 1, contributing to the processing of input information and learning long-range dependencies in sequential data. Specifically:
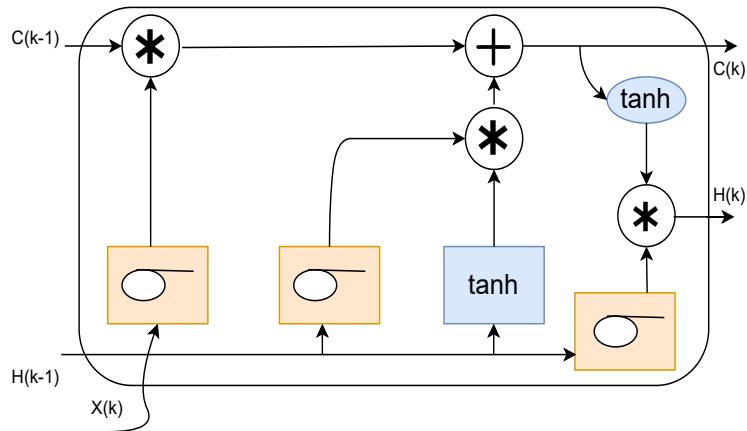
Figure 9. The Structure of LSTM Memory Unit.

The cell state, represented by the straight horizontal line on top of the network, undergoes updates at each time step, integrating the information and producing the final output. The combination of these sigmoid and tanh operations allows the LSTM cell to effectively capture and manage long-term dependencies in sequential data. In general, LSTM is defined by the following equations:

$$i(k) = a(W_i(H(k-1), X(k)) + b_i) \tag{4}$$

$$f(k) = a(W_f(H(k-1), X(k)) + b_f) \tag{5}$$

$$o(k) = a(W_o(H(k-1), X(k)) + b_o) \tag{6}$$

$$C\bar(k) = \tanh(W_c(H(k-1), X(k)) + b_c) \tag{7}$$

$$C(k) = f(k) * C(k-1) + i(k) * C\bar(k) \tag{8}$$

$$H(k) = o(k) * \tanh(C(k)) \tag{9}$$

The Equations (4)–(6) correspond to the input, forget, and output gates, respectively, in the LSTM cell. In these equations, $a(.)$ denotes the activation function for these gates, $X(k)$ represents the input vector at a time k, $C\bar(k)$ is the memory cell candidate, $C(k)$ denotes the memory cell state, and $H(k-1)$ denotes the hidden state. The symbols W and b represent the weight and bias parameters, respectively. Additionally, the symbol $*$ indicates element-wise multiplication. A memory cell candidate refers to the information that is proposed to be stored in the memory cell during the processing of input data. It undergoes a series of transformations before being potentially stored in the memory cell. The memory cell refers to the component responsible for storing and maintaining long-term dependencies in the sequential data.

To forecast a complete sequence of streamflow akin to the input, this model necessitates many-to-many structures of LSTM, which involves returning the representations of hidden states. In our examination, we employed the ReLU as the activation function for the hidden layers, as outlined below.

$$ReLU(z) = max(0, z) \tag{10}$$

The LSTM model was trained with a hidden layer consisting of 50 neuron units each and a dense layer corresponding to the look-ahead value for each sequence of input vs output, aimed at predicting the streamflow according to the output sequence. ReLU activation functions were incorporated into the hidden layer. The training process spanned 100 epochs with a batch size of 32. Batch size refers to the number of training examples used in one iteration of the optimization algorithm. It determines how many samples are processed before updating the model's parameters. The loss function utilized during training was Mean Square Error (MSE), calculated as the discrepancy between the observations and predictions and the optimizer is "Adam". The optimizer is a key element in training neural networks. It adjusts the model's parameters to minimize the loss function, improving overall performance. The most widely used optimizers are Adam, Stochastic Gradient Descent (SGD), and RMSprop. It is worth noting that the input shape was configured as (look-back value, *j*) to ensure compatibility with the sequence where *j* denotes the number of features used for training.

5.1.3. GRU

The GRU, or Gated Recurrent Unit, is a type of RNN architecture that offers a simpler structure compared to LSTM networks, featuring fewer gates for computational efficiency. It comprises key components such as the update gate, responsible for regulating the flow of new data into the current unit, and the reset gate, which determines which state variables should be retained or forgotten. The hidden state of the GRU represents the information propagated from the one-time step to the next. Additionally, the reset gate and update gate in GRU are designed to capture both short-term and long-term dependencies within sequential data. This makes GRU a versatile choice for various tasks involving sequential data processing.

$$z_t = \sigma(W_z.[H_{t-1}, X_t] + b_z) \tag{11}$$

$$r_t = \sigma(W_r.[H_{t-1}, X_t] + b_r) \tag{12}$$

$$\bar{H}_t = \tanh(W_h.[r_t * H_{t-1}, X_t] + b_h) \tag{13}$$

$$H_t = (1 - z_t) * H_{t-1} + z_t * \bar{H}_t \tag{14}$$

The Equations (11), (12) and (14) correspond to the update, rest, and output gates, respectively, in the GRU. $\bar{H}_t$ refers to the candidate activation vector. It is like a suggestion for the next memory state. It is calculated using input data and the previous memory state, offering a potential update for the new memory state. In these equations, σ represents the sigmoid function, X_t represents the input vector at time *t*, H_{t-1} denotes the previous memory state, and H_t denotes the final cell state. The symbols W and b represent the weight matrices and bias parameters. Additionally, the symbol $*$ indicates element-wise multiplication, the same as in the LSTM cell.

In our analysis, the GRU model was configured with the following parameters, manually set to optimize performance: GRU layer with 50 neuron units and ReLU activation function, as in Equation (10); input data shape set to look-back time steps and one feature, that is Flow; addition of a dense layer to the model, responsible for outputting predictions for the next look-ahead time steps; furthermore, the parameters for epochs, batch size, verbose, loss function, and optimizer were set to match those used for LSTM. This ensures consistency and facilitates comparison between the two models in terms of training and evaluation.

5.1.4. SARIMA

The SARIMA (Seasonal Autoregressive Integrated Moving Average) model is a time series forecasting method that incorporates both autoregressive (AR) and moving average (MA) components, as well as seasonal differencing. It is designed to analyze past data in order to make predictions about future values in a time series and is defined as:

$$SARIMA = c + \sum_{n=1}^{p} \alpha_n y_{t-n} + \sum_{n=1}^{q} \theta_n \epsilon_{t-n} + \sum_{n=1}^{P} \phi_n y_{t-mn} + \sum_{n=1}^{Q} \eta_n \epsilon_{t-mn} + \epsilon_t \quad (15)$$

Here, p is the number of AR terms, d is the degree of differencing indicating the number of times the series needs to be differenced to make it stationary, and q is the number of moving average terms. Other parameters P, D, Q, and m are seasonal autoregressive order, seasonal difference order, seasonal moving average order, and number of observations in a single seasonal cycle respectively. The values "d" and "D" are 0 for our SARIMA model. Setting, $d = D = 0$ means that no differencing is applied to the time series data, indicating that the data are already stationary and does not require differencing to make it stationary. We trained the SARIMA model with the optimal hyperparameters tailored for each combination of input and output sequences. The order and seasonal parameters of SARIMA are closely linked to the characteristics of the input and output data, necessitating training with the entire dataset. For the specific combination of input-output sequence, the best parameters determined were (p, d, q) for the order and (P, D, Q, m) for the seasonal order, with the values (3, 0, 2) and (2, 0, 1, 12) respectively.

5.2. Multivariate Time-Series Model Prediction

The correlation between temperature and streamflow is stronger than that of precipitation and snow water equivalent with streamflow, as depicted in Figure 7. But, the varying patterns in streamflow, as shown in Figures 4–6, correspond closely to changes in the other variables than the temperature. Thus, there appears to be a significant relationship between streamflow and meteorological parameters. In the multivariate time series analysis, the introduction of SNOTEL data such as snow water equivalent, average temperature, and precipitation has led to changes in the hyperparameter settings of the models. Here are the updated hyperparameters for each model:

1. RFR: In contrast to the univariate RFR model, hyperparameters were fine-tuned to optimize performance. Increased tuning, including: 500 estimators, Minimum samples split set to 10, Minimum samples leaf set to 5, Maximum depth set to 9, Utilization of bootstrap for resampling.
2. LSTM and GRU: No changes apart from modifying the input shape to (look-back size, 4) to accommodate the additional three attributes from the SNOTEL data in the input sequence.
3. SARIMA: Adjustments in order and seasonal order hyperparameters based on the combination of input-output sequences, as discussed in the subsection SARIMA of univariate time series analysis. No additional hyperparameters are introduced for this model.

These changes aim to enhance the models' ability to capture the relationships between the meteorological factors and streamflow, thereby improving the accuracy of the predictions in the multivariate setting.

5.3. Evaluation Metrics

In this section, we outline the metrics employed for comparing the machine learning models and conducting both univariate and multivariate analyses. The metrics utilized in our analysis include Mean Absolute Error (MAE), Root Mean Square Error (RMSE), Mean Absolute Percentage Error (MAPE), Symmetric Mean Absolute Percentage Error (SMAPE), and R-squared (R^2).

For clarification, let us consider y_i as the predicted value, x_i as the observed value, and n as the number of observations.

*   Mean Absolute Error (MAE): It calculates the average of the absolute differences between the predicted and actual values. It provides a measure of the average magni-

tude of errors in the predictions, without considering their direction and it is defined as follows:

$$MAE = \frac{1}{n} \sum_{i=1}^{n} |(y_i - x_i)| \tag{16}$$

- Root Mean Square Error (RMSE): It is a commonly used metric for evaluating the accuracy of a predictive model. RMSE is calculated by taking the square root of the average of the squared differences between predicted and actual values and defined as follows:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - x_i)^2} \tag{17}$$

- R-squared (R^2): It is a statistical measure that represents the proportion of the variance in the dependent variable (predicted values) that is explained by the independent variable (actual values). It is also known as the coefficient of determination. R^2 is defined as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - x_i)^2}{\sum_{i=1}^{n} (y_i - z_i)^2} \tag{18}$$

where z represents the mean of the actual values. The closer the value of R^2 to 1, the better the model's predictions compared to the actual values. However, it is essential to consider other evaluation metrics alongside (R^2) to gain a comprehensive understanding of the model's performance.

- Mean Absolute Percentage Error (MAPE): It is one of the most widely used metrics. It calculates the average percentage difference between predicted and actual values and is defined as:

$$MAPE = (\frac{1}{n}) \sum_{i=1}^{n} |\frac{y_i - x_i}{x_i}| * 100 \tag{19}$$

The Mean Absolute Percentage Error (MAPE) can encounter issues when the actual values in the dataset are close to or equal to zero, resulting in undefined or extremely large percentage errors. This problem can distort the evaluation of model performance, especially when dealing with datasets containing zero or near-zero values. Symmetric Mean Absolute Percentage Error (SMAPE) addresses the issue of asymmetric errors in MAPE by considering the average of the percentage error. It is also expressed as a percentage as follows:

$$SMAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - x_i|}{(|y_i| + |x_i|)/2} * 100 \tag{20}$$

6. Experimental Results

In this section, we will explore the detailed experimental outcomes of our analysis. This section is organized into subsections: (1) a comparison between Univariate and Multivariate Time Series, (2) a comparison of Four Machine Learning Models, (3) Sequence Analysis for RFR Model, and (4) Meteorological Factors Influence on Streamflow. We will delve into each of these areas in detail.

In our analysis, we explored various combinations of look-ahead and look-back periods to improve streamflow prediction accuracy. Through our investigation, we found that using an RFR model with a look-back period of 24 months and a look-ahead period of 12 months yielded the most favorable results. While our data allows for predicting streamflow up to 60 months ahead, we observed that the accuracy diminishes for longer forecast horizons. This is particularly noteworthy considering the potential variability introduced by climate changes over five years. Therefore, it is imperative to strike a balance between forecast horizon and accuracy when making long-term predictions. To evaluate

the performance of different machine learning models and their input-output sequence combinations, we employed five evaluation metrics as mentioned in the above section. These metrics provided comprehensive insights into the predictive capabilities of each model, enabling us to identify the most effective approach for streamflow prediction in our analysis.

6.1. Comparison of Univariate and Multivariate Time-Series

We conducted a thorough analysis by exploring multiple sequences of input and output combinations. A total of 15 possible combinations can be obtained for our data. The (input, output) sequences are organized into sets including pairs such as (12,12), (12,24), (12,36), (12,48), (12,60), (24,12), (24,24), (24,36), (24,48),(36,12), (36,24), (36,36), (48,12), (48,24), (60,12). The primary aim of this analysis was to assess the potential enhancement in RMSE achieved by integrating meteorological factors into the input sequence. Univariate models demonstrated superior performance initially. Upon comparison with multivariate models, the effectiveness of the univariate approach diminished. To visually depict these differences, we generated distribution plots for the RMSE values associated with each model, as shown in Figure 10.
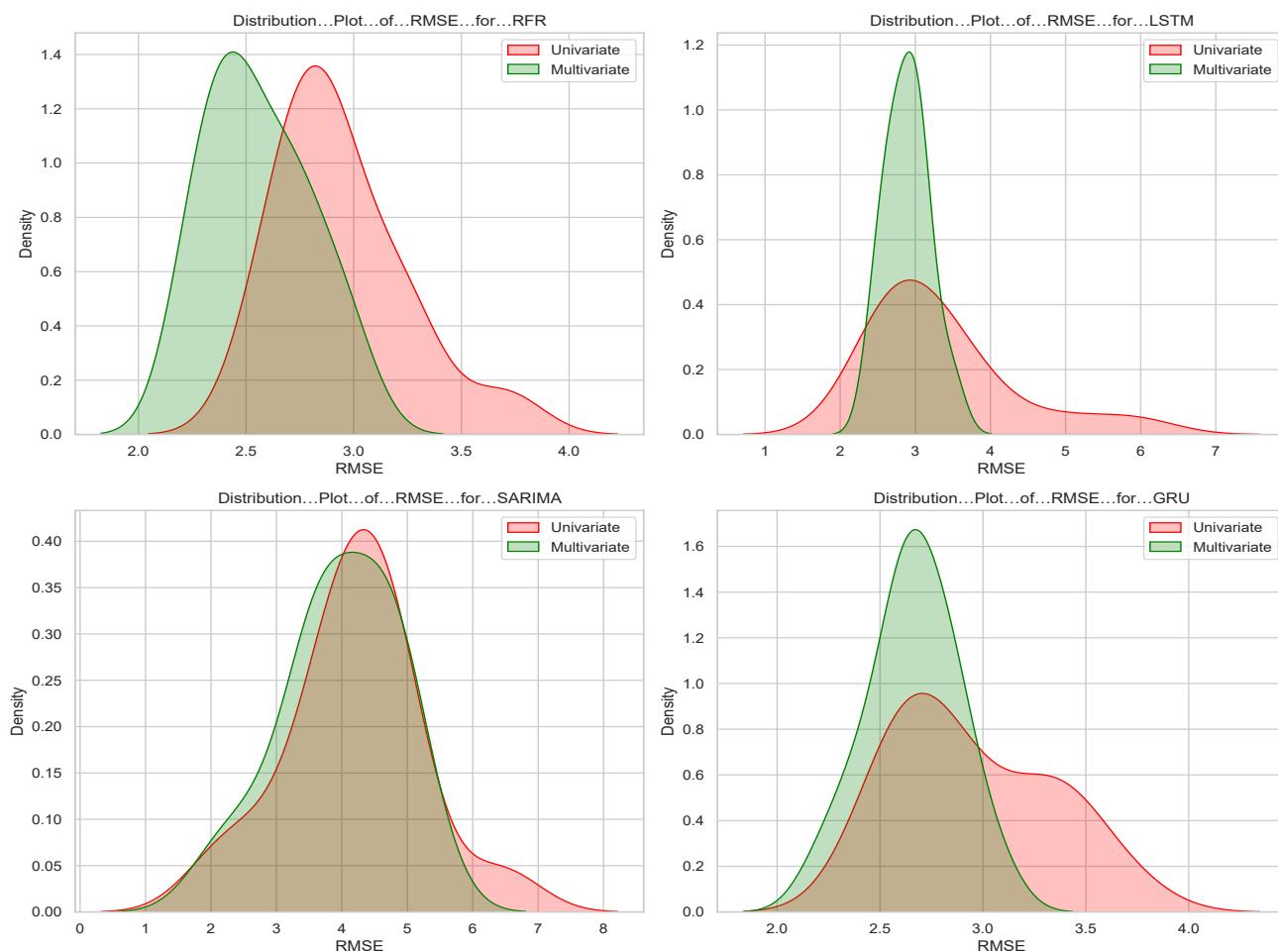
Figure 10. Comparison of RMSE Distribution between Univariate and Multivariate Models.

For multivariate LSTM models, we observed a significantly narrower range of RMSE values compared to their univariate counterparts. The width of the univariate RMSE range was approximately 6, whereas it reduced to nearly 2 for the multivariate models. Similarly, the density distribution for multivariate LSTM models exhibited a single peak, indicating a concentrated distribution of RMSE values within a specific range. Interestingly,

the univariate LSTM model achieved a lower minimum RMSE value than the multivariate LSTM model. For multivariate RFR models, although the minimum RMSE value decreased from around 2.1 to 1.6 compared to univariate models, the change in the range of values was not as significant as observed in RFR models. The density distribution showed a single peak for both univariate and multivariate models. Conversely, for SARIMA models, both univariate and multivariate distributions exhibited nearly identical minimum values, while the maximum value differed by almost 1.1. In univariate models, the maximum value was above 8, whereas in the multivariate model, it was 6.9. Additionally, in both models, the density showed peaks at a single value, which were almost the same. In GRU models, the range of RMSE values also decreased by approximately 0.75 in multivariate models compared to univariate ones. The density distribution showed a single peak for multivariate models, while for univariate models, there was a noticeable second peak. Multivariate GRU models exhibited the highest peak density, followed by RFR and LSTM. Univariate GRU models had a higher density for the second peak than the other.

The median R^2 plot comparing univariate and multivariate models demonstrated that all models performed better in terms of R^2 values when meteorological factors were included in the input sequence, Figure 11. Despite this difference not being so large, it still reflects a good improvement.
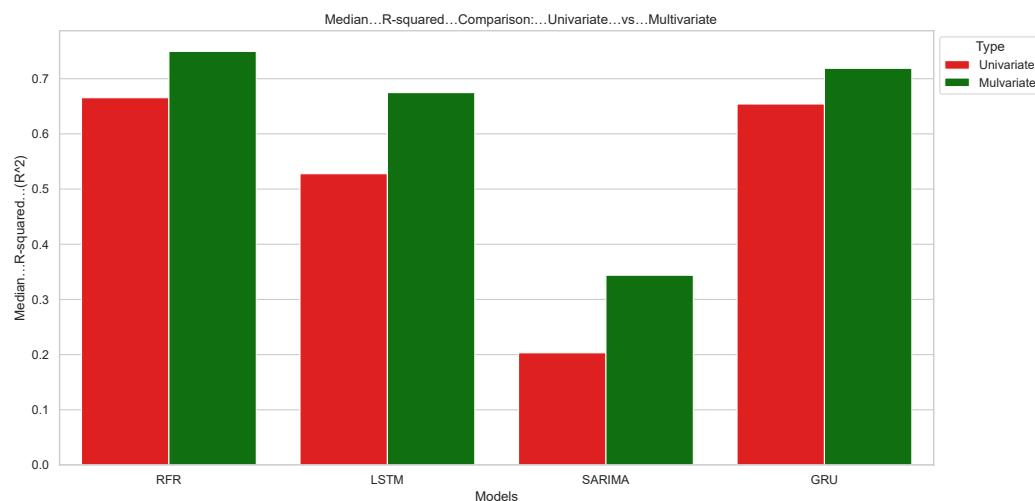
Figure 11. Comparison of Median R^2 between Univariate and Multivariate Time-Series Models.

The gap between the maximum R^2 values of univariate and multivariate models is greater in SARIMA and LSTM than in the RFR and GRU. However, it is important to highlight that there were instances in the univariate SARIMA approach where the R^2 values turned negative, indicating poor model performance, Figure 12. While there was one combination where the SARIMA model outperformed RFR, LSTM, and GRU in terms of R^2 values, this superiority was not consistent across other combinations. Adjusting hyperparameters did not lead to improvement in cases where the R^2 values were found to be negative. This observation underscores the SARIMA model's reliance on streamflow data, with limited impact from additional meteorological factors incorporated in the multivariate approach.

Based on Figures 10 and 11, we found that integrating snow water equivalent, temperature, and precipitation data from the Upper Colorado River Basin into streamflow prediction models leads to a reduction in RMSE values and enhanced R^2 values.
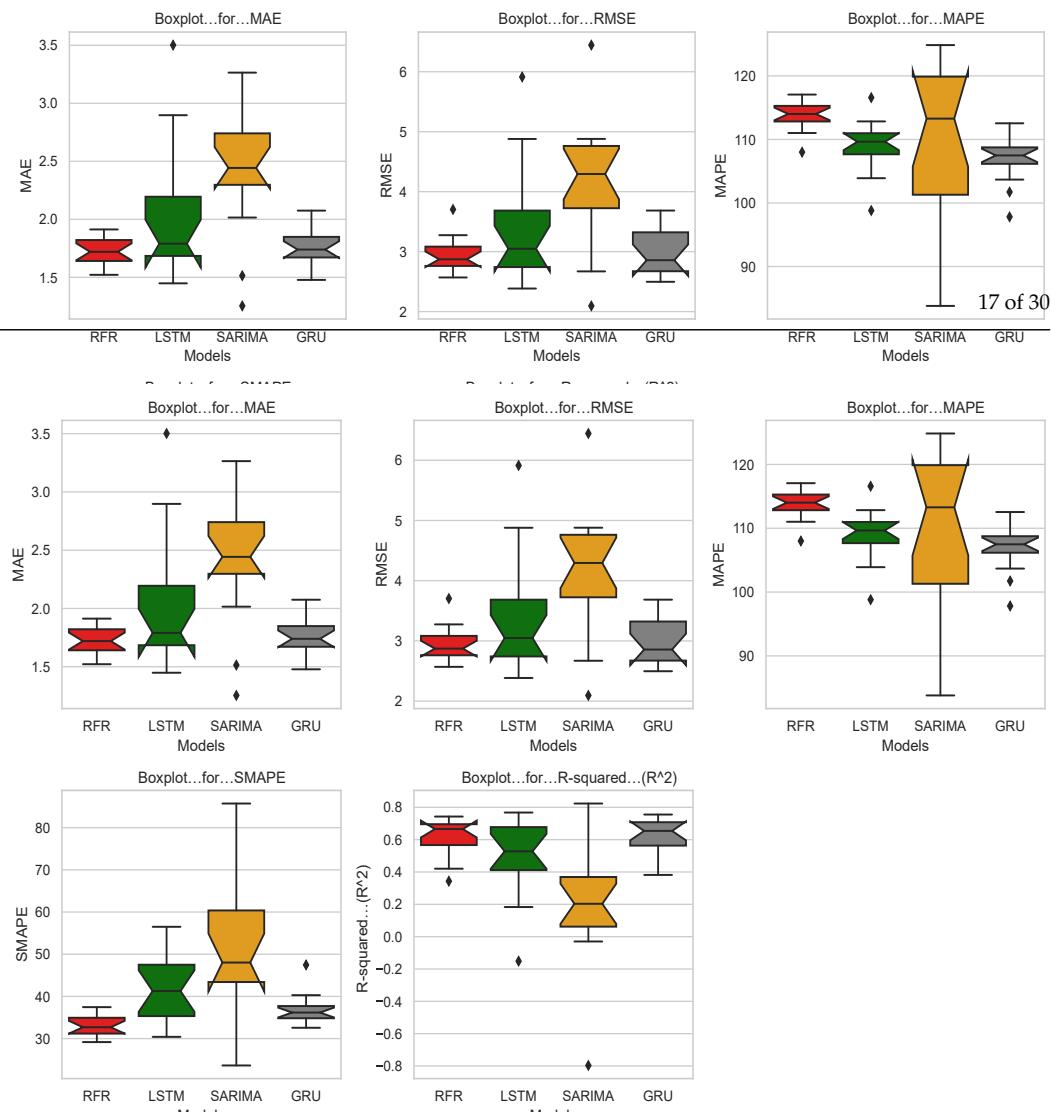
Figure 12. MAE, RMSE, SMAPE, MAPE, and R^2 Results of Univariate Time-Series ML Models.

with the outliers. In the presence of an outlier in the RFR RMSE plot, the median values of RMSE for both RFR and GRU are nearly identical.

Figure 13. RMSE Distribution Plot for Univariate Time-Series Models.

From Figure 13, we can see that the MAE values for both RFR and GRU models are quite similar, at 1.72 and 1.74 respectively, with an interquartile range of 0.18 for both. On the other hand, LSTM and SARIMA models have larger interquartile ranges of 0.44 and 0.51 respectively, indicating more variability in their performance. In terms of RMSE, the median values for RFR and GRU are also close, at 2.86 and 2.87 respectively, with LSTM

From Figure 13, we can see that the MAE values for both RFR and GRU models are quite similar, at 1.72 and 1.74, respectively, with an interquartile range of 0.18 for both. On the other hand, LSTM and SARIMA models have larger interquartile ranges of 0.44 and 0.51, respectively, indicating more variability in their performance. In terms of RMSE, the median values for RFR and GRU are also close, at 2.86 and 2.87, respectively, with LSTM slightly higher at 3.05. When looking at MAPE, GRU performs the best with a value of 109, followed by LSTM with 107.67, and RFR closely resembling SARIMA at 113.3 and 114, respectively. SARIMA exhibits a minimum value of around 82 in the plot, which is notably distant from its median. The same trend is observed in SMAPE, with a minimum of around 22 and a median of 48 for SARIMA, while RFR has a median of 32.7 with an interquartile range of 3.75, and GRU with a smaller interquartile range of 2.89 and a median of 36.17. Consistently, RFR shows the highest median R^2 value at 0.67, followed closely by GRU at 0.65, while LSTM trails at 0.41 and SARIMA with the lowest median R^2 value of 0.2. These observations suggest that in the univariate model, RFR outperforms the other models.

In the next step, we plotted the RMSE distribution for multivariate models, Figure 14. It is observed that the RFR RMSE values have an overall lower magnitude compared to others. However, the spread of GRU is less dense, mainly concentrated in a certain range. The stretch of GRU lies between that of RFR and LSTM, where GRU is less stretched compared to its distribution in the univariate approach (as depicted in Figure 13). Similar to the univariate distribution, SARIMA's distribution is stretched, with density located differently from the others. Comparatively, RFR, LSTM, and GRU exhibit narrower RMSE error distributions compared to their univariate model.
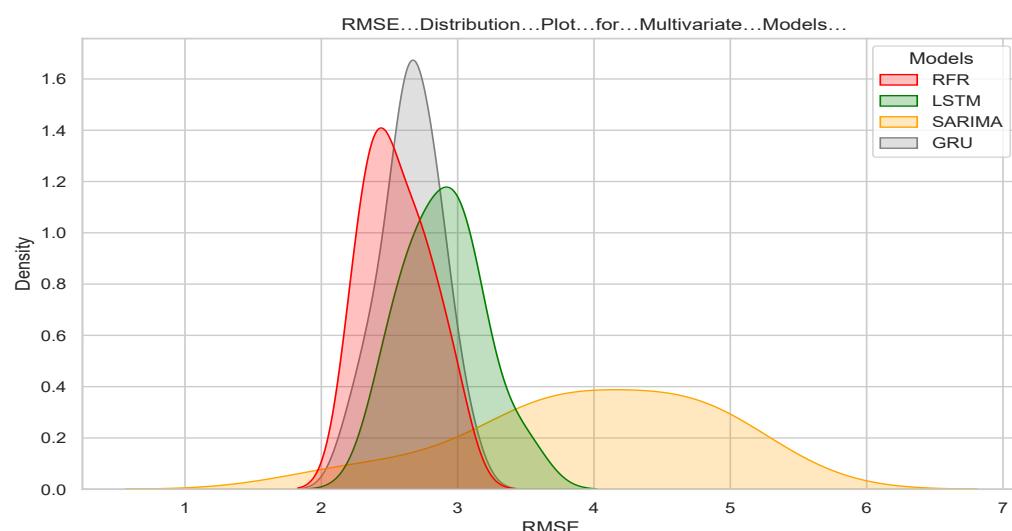
Figure 14. RMSE Distribution of Multivariate Time-Series Models.

Upon examining the box plots of multivariate models in Figure 15, we did not notice any significant increase in variability for the SARIMA model compared to its univariate counterpart. RFR showed improvements across all metrics in the multivariate approach, with median values of MAE, RMSE, MAPE, SMAPE, and R^2 being around 1.51, 2.51, 112, 30, and 0.78, respectively. In contrast, in the univariate approach, these values were around 1.74, 2.9, 113, 32, and 0.66, respectively. LSTM's MAE median remained the same as its univariate but exhibited a narrower interquartile range and better R^2 values. In addition to RFR, GRU performed well in the multivariate approach, with improved medians of MAE from 1.75 to 1.6, RMSE from 2.9 to 2.6, and R^2 from 0.65 to 0.7, and overall, displayed a narrower interquartile range for these metrics. It is worth noting that RFR, LSTM, and GRU each had outliers for the R^2 value. Indeed, across both the univariate and multivariate approaches, RFR consistently outperformed all other models in terms of various evaluation metrics.

Figure 14. RMSE Distribution of Multivariate Time-Series Models.
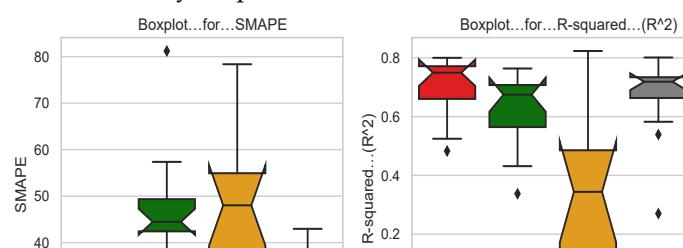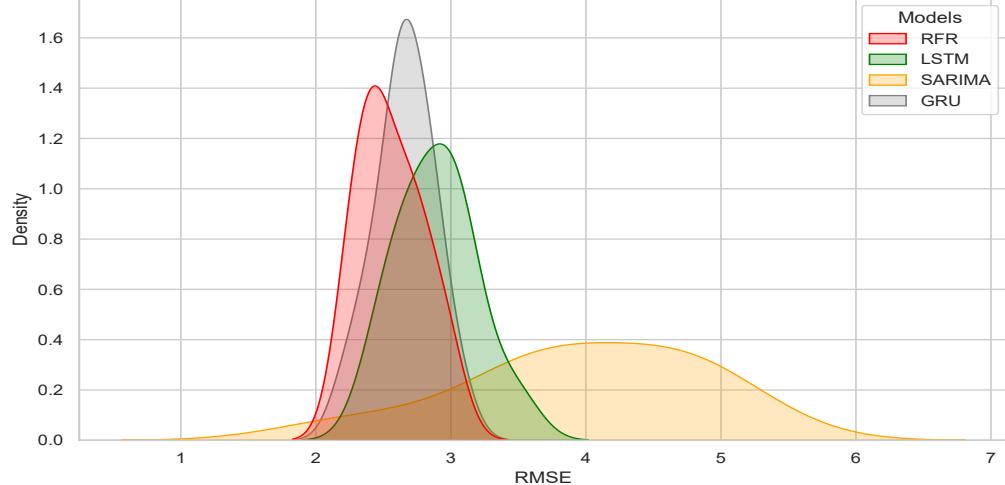
values for the remaining models are depicted in Table 1. A comparison of the predicted and observed USBR streamflow of these models is shown in Figure 16.
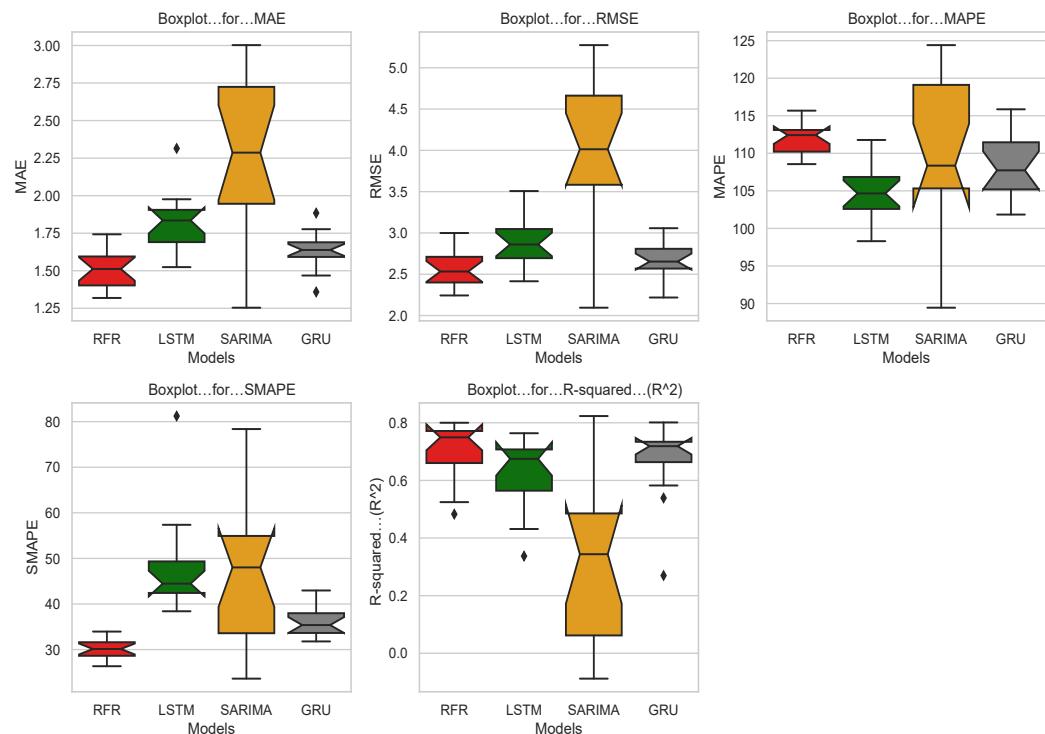
Figure 15. RMSE, SMAPE, MAPE, and R^2 Results of Multivariate Time-Series ML Models.

Our analysis revealed that the predicted values for the period from May 2015 to August 2015 exhibited greater variability for the 12-12 combination compared to the 24-12 combination when compared against the observed values. However, the streamflow values remained almost identical for the subsequent seven months, making the error values for the 12-12 combination similar to those for the 24-12 combination. Upon careful examination of both the graphs and the error values, we determined that the 24-12 combination of RFR, with MAE of 1.3, RMSE of 2.2, MAPE of 109, SMAPE of 27.3, and R^2 of 0.8, represented the superior model for this dataset. Figure 17 illustrates the predicted streamflow using the optimal 24-12 input-output combination compared against the observed values for the discussed models.
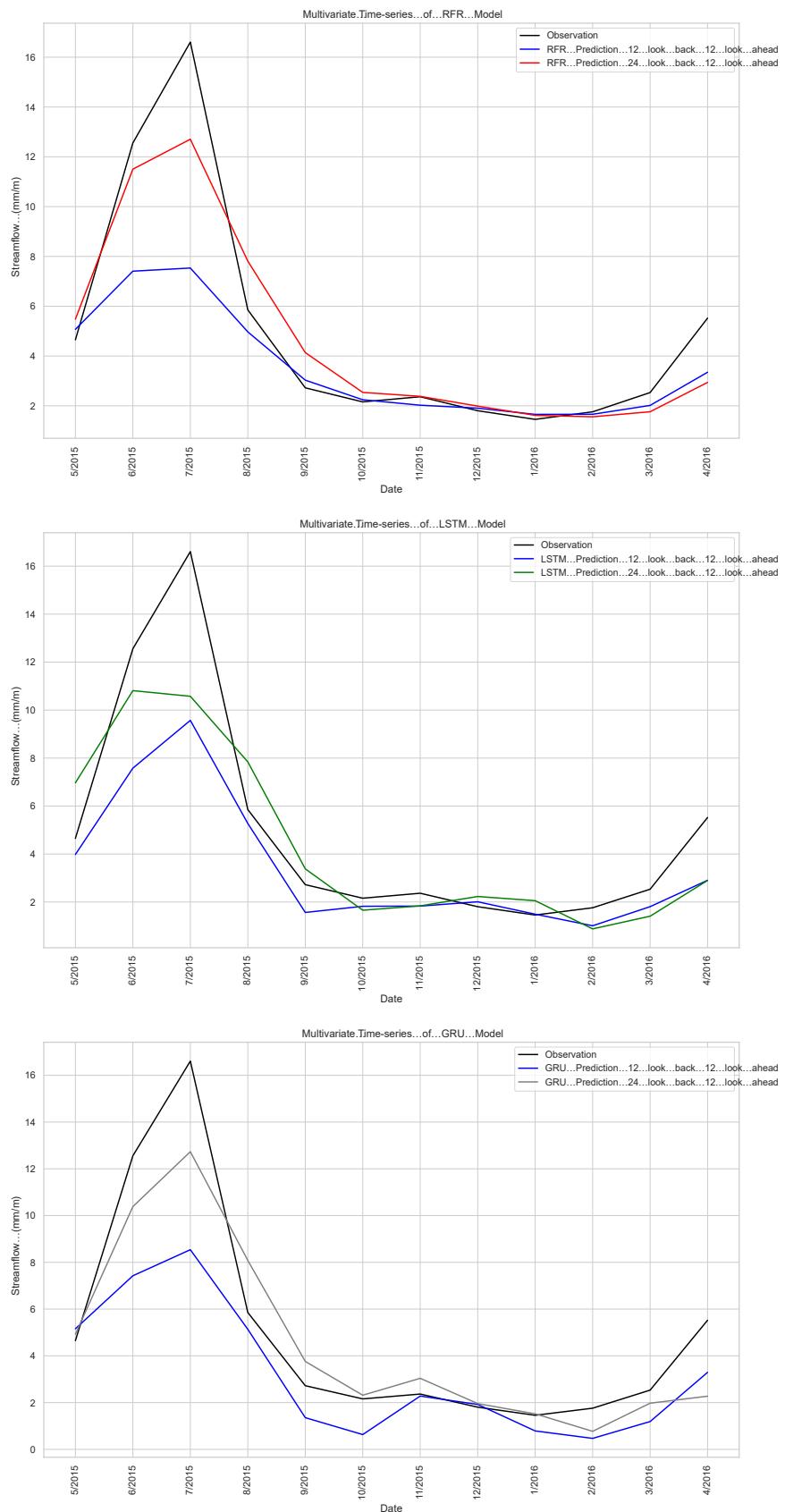
Figure 16. Comparison of Predicted and Observed USBR Streamflow Using Multivariate Time-Series Models (RFR, LSTM, GRU) with Input-Output Combinations of 12-12 and 24-12, for the Period from May 2015 to April 2016 over the Test Set.
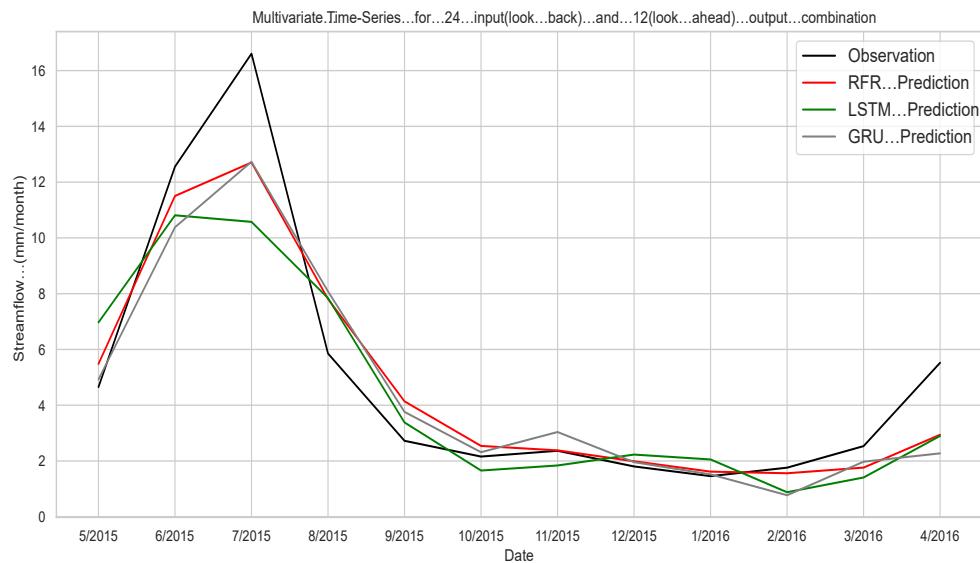
Figure 17. Comparison of Predicted vs. Observed Streamflow from Multivariate Time-Series Models (RFR, LSTM, GRU) over the Test Set (May 2015–April 2016).

Table 1. Model Performance Metrics.

Models	Metrics				
	MAE	**RMSE**	**MAPE**	**SMAPE**	**R-Squared**
Univariate					
RFR	1.52	2.57	111.01	29.17	0.74
LSTM	2.14	3.63	108.62	47.50	0.49
GRU	1.48	2.50	103.68	32.83	0.75
Multivariate					
RFR	1.34	2.25	109.34	27.33	0.80
LSTM	1.60	2.58	99.98	43.30	0.74
GRU	1.50	2.40	105.28	32.80	0.77

6.3. Sequence Analysis for RFR Model

In addition to the previous experiments, we conducted a comparison of different input (look back) and output (look ahead) sequence combinations for the best model, RFR, both in univariate and multivariate scenarios. This comparison aimed to observe the variations in prediction quality resulting from different combinations of look ahead and look back sequences. We considered a range of look back and look ahead span windows, including 12, 24, 36, 48, and 60 months. As outlined in the Methodology section, we explored all possible combinations where the sum of the sequence lengths equaled 72 months (6 years), matching the length of the testing set. Assigning 80% of a dataset to training is commonly regarded as a best practice in many articles [68]. This allocation ensures an ample amount of data for both modeling and testing. Utilizing a smaller portion of the data for training can lead to less effective models [69,70]. The main objective of this experiment is to observe how the predictions or errors vary with changes in the combinations of input and output sequences.

Figure 18 shows heat maps of evaluation metrics for the univariate time-series RFR model. From the heatmap, it is evident that increasing the output sequence length with a particular input sequence tends to result in higher errors across all metrics. However, this trend is not consistent for the input sequence length of 12 months, where the values do not follow a clear pattern. Notably, interesting results are observed for the input sequences predicting 12-month and 48-month outputs, yielding similar MAE (1.6) and R^2 (0.7) metrics, although RMSE (2.76 and 2.68) and SMAPE (30.67 and 31.51) values differ

6.3. Sequence Analysis for RFR Model

In addition to the previous experiments, we conducted a comparison of different input (look back) and output (look ahead) sequence combinations for the best model, RFR, both in univariate and multivariate scenarios. This comparison aimed to observe the variations in prediction quality resulting from different combinations of look ahead and look back sequences. We considered a range of look back and look ahead span windows, including 12, 24, 36, 48, and 60 months. As outlined in the Methodology section, we explored all possible combinations where the sum of the sequence lengths equaled 72 months (6 years), matching the length of the testing set. Assigning 80% dataset to training is commonly regarded as a best practice in many articles [60]. This allocation ensures an ample amount of data for both modeling and testing. Utilizing a smaller portion of the data for training can lead to less effective models [61,62]. The main objective of this experiment is to observe

slightly. The prediction of a 24-month output sequence with a 12-month input sequence outperforms other combinations with this input, demonstrating the lowest MAE (1.53), RMSE (2.61), and highest R^2 (0.72) values. Analyzing specific output sequence lengths with increasing input sequence lengths reveals that the model's performance does not significantly improve, except for the 24-month input and 12-month output combination, which stands out as the optimal configuration for the RFR model. This combination yields ... (0.74)

MAPE

Output↓ / Input→	12	24	36	48	60
60	115.00	0.00	0.00	0.00	0.00
48	113.00	113.00	0.00	0.00	0.00
36	114.00	115.00	117.00	0.00	0.00
24	113.00	114.00	117.00	116.00	0.00
12	112.00	111.00	114.00	115.00	108.00

SMAPE

Output↓ / Input→	12	24	36	48	60
60	34.70	0.00	0.00	0.00	0.00
48	31.51	33.97	0.00	0.00	0.00
36	32.44	34.41	37.45	0.00	0.00
24	30.08	32.69	35.45	35.58	0.00
12	30.67	29.17	32.55	35.16	30.86

R-squared (R^2)

Output↓ / Input→	12	24	36	48	60
60	0.52	0.00	0.00	0.00	0.00
48	0.70	0.42	0.00	0.00	0.00
36	0.68	0.67	0.34	0.00	0.00
24	0.72	0.70	0.66	0.45	0.00
12	0.70	0.74	0.68	0.63	0.61

MAE

Output↓ / Input→	12	24	36	48	60
60	1.72	0.00	0.00	0.00	0.00
48	1.61	1.78	0.00	0.00	0.00
36	1.72	1.75	1.91	0.00	0.00
24	1.53	1.66	1.78	1.86	0.00
12	1.63	1.52	1.72	1.90	1.91

RMSE

Output↓ / Input→	12	24	36	48	60
60	2.79	0.00	0.00	0.00	0.00
48	2.68	2.95	0.00	0.00	0.00
36	2.90	2.87	3.18	0.00	0.00
24	2.61	2.77	2.98	3.24	0.00
12	2.76	2.57	2.86	3.27	3.70

(x-axis: Input...Sequence; y-axis: Output...Sequence)

Figure 18. Heatmap Visualization of Evaluation Metrics (MAPE, SMAPE, R^2, MAE, and RMSE) for Univariate RFR Model.

Figure 18 shows the visual representation of all the results. Figure 19, univariate RFR model. From the look-back of 12, it is evident that the predicted values with the shorter length values in the particular input sequence tend to result with higher errors. Subsequently, with a look-back of 24 months, the predicted values become closer to the observed values. The patterns for 48 months and 12 months follow a similar trajectory in Figure 19 as well. However, with a look-back of 36 months, the predictions deviate significantly from the observed values, showcasing a different graph compared to other combinations.
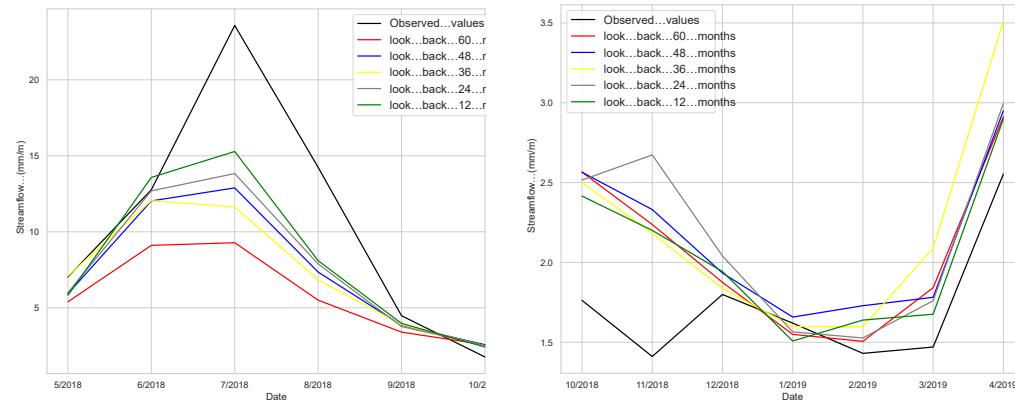
Figure 19. Predictions of Univariate Time-Series RFR Models with Different Look-back Span Windows against Ground Truth (May 2018–April 2019).

Next, we plotted the error heatmaps of the multivariate RFR models to analyze variations in the results, similar to the univariate observations, and to compare these results with the univariate observations. Figure 20, depicting the heat map of evaluation metrics for the multivariate RFR models, the impact of changing input and output sequences on error values was examined.

although RMSE (2.76 and 2.68) and SMAPE (30.67 and 31.51) values differ slightly. The prediction of a 24-month output sequence with a 12-month input sequence outperforms other combinations with this input, demonstrating the lowest MAE (1.53), RMSE (2.61), and highest R^2 (0.72) values. Analyzing specific output sequence lengths with increasing input sequence lengths reveals that the model's performance does not significantly improve, except for the 24-month input and 12-month output combination, which stands out as the optimal configuration for the RFR model. This combination yields the lowest MAE (1.52), RMSE (2.57), MAPE (111), SMAPE (26:17) and highest R^2 (0.74) by taking in a 24-month input sequence.

Figure 19. Predictions of Univariate Time-Series RFR Models with Different Look-back Span Windows against Ground Truth (May 2018 - April 2019).

The visual representation of the results, Figure 19 confirms this observation. Initially, with a look-back of 12 months, the predicted values align with the observed values in the first 5 months when compared with others. Subsequently, with a look-back of 24 months, the predicted values become closer to the observed values. The patterns for 48 months and 12 months follow a similar trajectory in Figure 19 as well. However, with a look-back of 48 months, the predicted values appear to deviate further from the observed values compared to the 24-month look-back period, although it achieves a higher R^2 value of 0.8.

The trend continues with the R^2 plot, where the 60-month input sequence exhibits a higher value of 0.77 compared to 0.75 for both 48 and 60 months. The highest R^2 value is observed for predicting 12 months of data with 12 and 24 months of input, and the other metrics for these two combinations also show slight differences. From Figure 21, it is apparent that the predicted flows for 36 and 60 months follow a similar trajectory for 9 out of the 12 months, yet there is not a clear correlation between them in terms of the metrics. Similarly, there seems to be a relationship between 48 and 12 months. Notably, the 12-month input sequence appears to deviate further from the observed values compared to the 24-month look-back period, although it achieves a higher R^2 value of 0.8. Reducing the magnitude of the graph

MAPE

Output / Input	12	24	36	48	60
60	111.00	0.00	0.00	0.00	0.00
48	110.00	111.00	0.00	0.00	0.00
36	113.00	113.00	115.00	0.00	0.00
24	110.00	112.00	116.00	115.00	0.00
12	109.00	109.00	113.00	113.00	110.00

SMAPE

Output / Input	12	24	36	48	60
60	29.73	0.00	0.00	0.00	0.00
48	29.12	30.58	0.00	0.00	0.00
36	30.13	32.13	33.94	0.00	0.00
24	28.17	30.71	33.83	33.95	0.00
12	26.35	27.33	30.13	31.09	28.11

R-squared (R^2)

Output / Input	12	24	36	48	60
60	0.59	0.00	0.00	0.00	0.00
48	0.75	0.52	0.00	0.00	0.00
36	0.77	0.74	0.48	0.00	0.00
24	0.79	0.77	0.73	0.59	0.00
12	0.80	0.80	0.77	0.75	0.75

MAE

Output / Input	12	24	36	48	60
60	1.53	0.00	0.00	0.00	0.00
48	1.50	1.62	0.00	0.00	0.00
36	1.40	1.57	1.71	0.00	0.00
24	1.38	1.48	1.62	1.74	0.00
12	1.32	1.34	1.40	1.51	1.56

RMSE

Output / Input	12	24	36	48	60
60	2.54	0.00	0.00	0.00	0.00
48	2.46	2.71	0.00	0.00	0.00
36	2.40	2.53	2.81	0.00	0.00
24	2.29	2.43	2.68	3.00	0.00
12	2.24	2.25	2.40	2.72	2.92

Input...Sequence

Figure 20. Heatmap Visualization of Evaluation Metrics (MAPE, SMAPE, R^2, MAE, and RMSE) for Multivariate RFR Model.

Next, we plotted the error heatmaps of the multivariate RFR models to analyze variations in the results, similar to the univariate observations, and to compare these results for both univariate and multivariate RFR models. Based on these observations and as discussed in the above section (Figure 14), predicting 12 months with a 24-month input is considered the optimal RFR model.

The experiment concludes that for this dataset, increasing the sequence values of the input does not improve performance beyond a certain point, which is observed to be 24 for both univariate and multivariate RFR models.
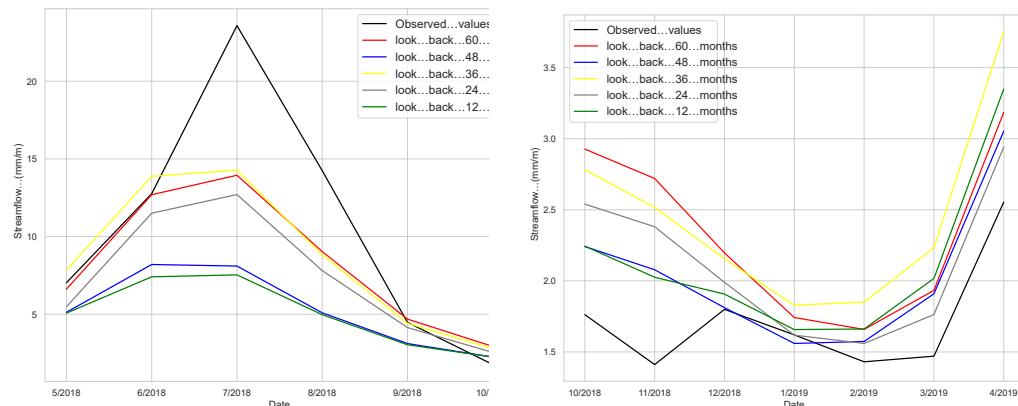
Figure 21. Predictions of Multivariate Time-Series RFR Models with Different Look-back Span Windows against Ground Truth (May 2018–April 2019).

6.4. Ablation Analysis

In this section, we investigated the influence of meteorological factors on streamflow. To achieve this, we systematically included various combinations of these factors alongside streamflow in the input sequences to predict the output. Each combination of input and output sequences resulted in seven different feature combinations, as the streamflow remained constant while the meteorological factors varied. Subsequently, we trained machine learning models using only meteorological factors as features to predict streamflow values, excluding the streamflow data themselves. This approach aimed to discern the models' dependency on these factors in predicting streamflow. Better results are obtained when streamflow is included in the features alongside meteorological factors. Across all sequences, we examined both scenarios and identified the top three feature combinations for each sequence. It is important to note that there was no consistent top feature combination for all the sequences. These observations yielded intriguing results. We analyzed the top three feature combinations for each sequence based on their RMSE and R^2 values, considering their importance in evaluating the model.

In Figure 22, "S" represents Snow Water Equivalent, "P" stands for Precipitation Accumulation, and "T" denotes Average Temperature. The columns labeled as RMSE_1_F indicate the first RMSE value with historic streamflow included in the features, while RMSE_1 represents the first RMSE value without historic streamflow in the features. Similarly, the remaining labels follow this pattern, where the suffix "F" denotes the presence of historic streamflow in the features, while the absence of this suffix indicates the exclusion of historic streamflow. It is noteworthy that both "P" and "SP" combinations have the highest count for RMSE and R^2 values, with only one count for all combinations, "SPT". Conversely, the "ST" combination has the lowest count in both scenarios. "SPT" has a good count, but "SP" and "P" are denser in their distribution. Upon tallying all values, we found that the highest count is for "SPT" (45), followed by "SP" (44), and then "P" (39), while the rest have lower counts. To gain a deeper understanding of the important combinations, we plotted the feature importance for the RFR model, Figure 23. From this plot, it is evident that temperature is a highly important feature for predicting streamflow, despite having only 13 counts (alone) in the top three combinations. Precipitation and snow water equivalents exhibit almost the same importance, reflected by their combination count of 44.

Based on all the observations from our analysis, we conclude that all three factors—precipitation, snow water equivalent, and average temperature—need to be included in the features along with historic streamflow to obtain unbiased and optimal results.

Count...of alues
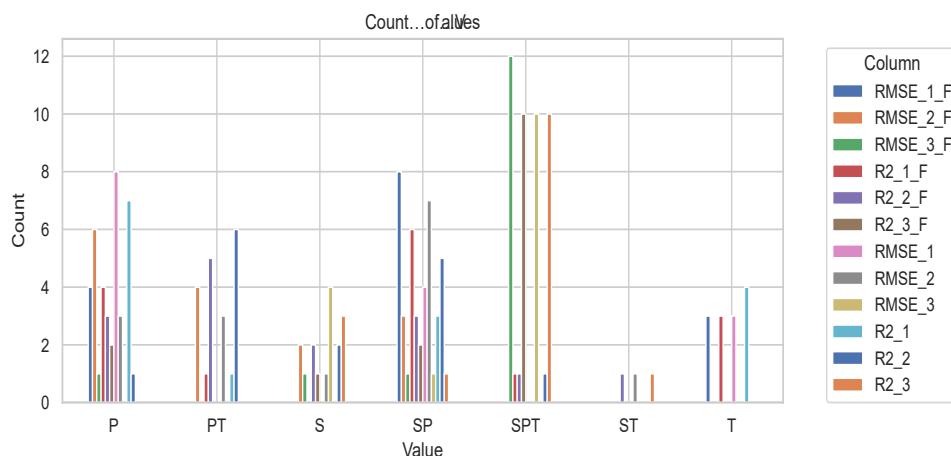
Figure 22. Top Features' Combinations Count.

Feature...Importance...for...RFR...Model
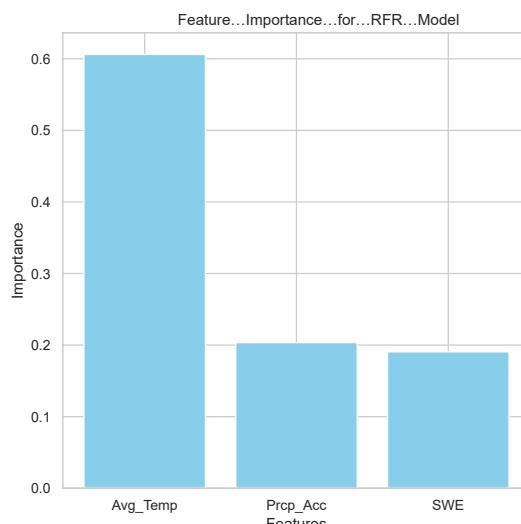
Figure 23. Feature Importance of Meteorological Factors.

7. Discussion

In our study, we found that the RFR model achieved optimal results in both univariate and multivariate scenarios when all meteorological factors were included to predict monthly streamflow data. Our findings are consistent with previous studies by Hosseinzadeh et al. [21] and Xu et al. [35] which show that multivariate models perform better than the univariate models. Our results also agree with Zhao et al. research [36] that including surface temperature can improve predictions. Additionally, Figure 23 highlights how important temperature is as a contributing factor in our models. Following RFR, the GRU model performed better than the LSTM model. LSTM has been reported to perform well in predicting streamflow in snowmelt regions [63] and has outperformed models like Support Vector Regression (SVR) [64] in previous studies. Some studies stated that LSTM and GRU can achieve good results for predicting streamflow [65] but we observed that standalone LSTM did not yield satisfactory results in both univariate and multivariate scenarios for our dataset. The effectiveness of the LSTM model depends on several factors, including the volume of data available, the nature of feature relationships (whether linear or not), the optimization of hyperparameters, and the complexity of the model architecture. In our study, we trained the model using monthly streamflow data spanning 24 years. However, our analysis revealed that the predicted feature did not exhibit a strong linear correlation with the input features, Figure 7, likely due to the basin's extensive geographical coverage and varying altitudes. These factors likely contributed to the model's performance. Conversely, the results obtained with the GRU model were comparable to those of the RFR model.

The superior performance of RFR compared to other ML models may be attributed to its architecture, which helps in mitigating overfitting. However, it is worth noting that there are studies where RFR has outperformed models like LSTM [66], SVM and Neural Networks [67], while in other cases, the results have been the opposite [68]. Among the ML

varying altitudes. These factors likely contributed to the model's performance. Conversely, the results obtained with the GRU model were comparable to those of the RFR model.

The superior performance of RFR compared to other ML models may be attributed to its architecture, which helps in mitigating overfitting. However, it is worth noting that there are studies where RFR has outperformed models like LSTM [74], SVM and Neural Networks [75], while in other cases, the results have been the opposite [76]. Among the ML models tested, SARIMA exhibited the least performance and required considerable time to adjust hyperparameters for satisfactory results. In time series-based methods, SARIMA has been shown to outperform other models such as ARIMA, and ARMA [77], and it even performed better than LSTM in monthly streamflow predictions [21]. However, in our analysis using 30 years of monthly data, LSTM performed better than SARIMA. Recent studies have highlighted SARIMA's reliability for longer lead-time forecasts, although its effectiveness may diminish when trained on short observation periods due to overfitting issues [32,42] and they have demonstrated good performance in drought conditions [78], which could explain their underperformance in this study. However, the performance of SARIMA models could potentially be improved with access to larger datasets.

To evaluate the generalizability of our best-performing model, RFR, we assessed its performance across three distinct locations in the Upper Colorado River Basin: Maybell, along the Yampa River in Colorado; Archuleta, on the San Juan River in New Mexico; and Greenwood, along the Colorado River in Colorado. These sites were chosen for their diverse geographical positions within the UCRB and their significant influence on flow dynamics at Lees Ferry, as depicted in Figure 2. Streamflow at these locations is notably lower than at Lees Ferry, providing indications of the model's adaptability. Our evaluation revealed consistently smaller errors at these locations compared to Lees Ferry. In both univariate and multivariate scenarios, Maybell consistently had the lowest MAE values, averaging 0.12 and 0.21, respectively, while Lees Ferry had much higher MAE values, reaching 1.74 for univariate and 1.5 for multivariate. Similarly, in terms of RMSE for both univariate and multivariate scenarios, Lees Ferry showed higher values (2.9 and 2.51) compared to Maybell (0.22, 0.21) and Archuleta (0.39, 0.38, respectively). The evaluation metrics for Archuleta and Glenwood exhibit similar values. The RFR model performed better at alternative locations than at Lees Ferry, with Maybell achieving the highest R^2 values of 0.85 (single-variable) and 0.868 (multiple-variable). These results are depicted in Figure 24. These findings underscore the model's robust performance in regions with lower streamflow. However, it is important to note that our analysis did not address extreme events such as floods or droughts.
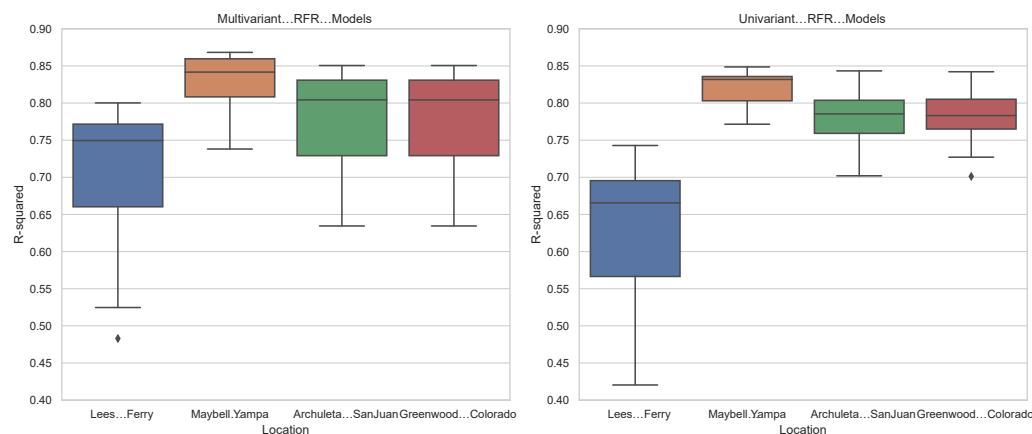

Figure 24. Boxplot of R^2 for RFR Model Across Different Streamflows.

8. Conclusions

In this study, we used machine learning models to forecast streamflow in the Upper Colorado River Basin (UCRB), focusing on Lees Ferry, a key point for water flow into

the Lower Colorado River Basin (LCRB). Data from 1991 to 2020 were obtained from the United States Bureau of Reclamation (USBR) for streamflow and from SNOTEL weather stations for meteorological data—snow water equivalent, precipitation, and temperature. We trained four machine learning models (RFR, LSTM, GRU, and SARIMA) using monthly data. Initially, we trained the models on past streamflow data (univariate approach) and then incorporated meteorological factors (multivariate approach). Evaluation with metrics like RMSE, R^2, and MAE showed that the multivariate models outperformed the univariate ones, indicating the positive influence of meteorological factors on streamflow prediction. The RFR model yielded the best results overall, and the GRU model outperformed LSTM. We also explored the impact of meteorological factors and input-output combination sequences on prediction. A combination of 24 months of input data predicting 12 months of output data yielded the best results.

Based on our investigation, this study represents the first attempt to examine the impact of meteorological factors on streamflow prediction while incorporating sliding window concepts for both look-back and look-ahead sequences. These findings offer opportunities to enhance ML model performance by incorporating additional factors and data. Our study has limitations, such as not extensively discussing certain aspects like the influence of temperature and not exploring hybrid models. Based on these findings, there is an opportunity to extend this study by including more meteorological variables, enlarging the dataset, and investigating alternative graph-structured modeling techniques like Graph Neural Networks. This could result in stronger conclusions and improved predictive precision. Examining model performance within defined radii around chosen SNOTEL stations as well as using diverse spatial shapes instead of all SNOTEL data, could provide valuable insights into localized streamflow behavior.

Author Contributions: S.T. implemented, analyzed, and wrote the paper. A.N. and S.F.B. supervised the research, contributed ideas during the interpretation of results, and reviewed the paper. S.M.H. contributed with ideas and reviewed the paper. P.H. contributed with writing—review and editing. All authors have read and agreed to the published version of the manuscript.

Data Availability Statement: The data presented in this study are openly available on the United States Bureau of Reclamation (USBR) website at https://www.usbr.gov/lc/region/g4000/NaturalFlow/current.html (accessed on 19 September 2023).

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Willett, S.D.; McCoy, S.W.; Perron, J.T.; Goren, L.; Chen, C.Y. Dynamic Reorganization of River Basins. *Science* **2014**, *343*, 1248765. [CrossRef]
2. Chung, M.G.; Frank, K.A.; Pokhrel, Y.; Dietz, T.; Liu, J. Natural Infrastructure in sustaining global urban freshwater ecosystem services. *Nat. Sustain.* **2021**, *4*, 1068–1075. [CrossRef]
3. Baker, S.A.; Wood, A.W.; Rajagopalan, B.; Prairie, J.; Jerla, C.; Zagona, E.; Butler, R.A.; Smith, R. The Colorado River Basin Operational Prediction Testbed: A Framework for Evaluating Streamflow Forecasts and Reservoir Operations. *JAWRA J. Am. Water Resour. Assoc.* **2022**, *58*, 690–708. [CrossRef]
4. Turner, S.W.D.; Bennett, J.C.; Robertson, D.E.; Galelli, S. Complex relationship between seasonal streamflow forecast skill and value in reservoir operations. *Hydrol. Earth Syst. Sci.* **2017**, *21*, 4841–4859. [CrossRef]
5. Wang, W.-C.; Chau, K.-W.; Cheng, C.-T.; Qiu, L. A Comparison of Performance of Several Artificial Intelligence Methods for Forecasting Monthly Discharge Time Series. *J. Hydrol.* **2009**, *374*, 294–306. [CrossRef]
6. Guo, J.; Zhou, J.; Qin, H.; Zou, Q.; Li, Q. Monthly streamflow forecasting based on improved support vector machine model. *Expert Syst. Appl.* **2011**, *38*, 13073–13081. [CrossRef]
7. Young, C.-C.; Liu, W.-C. Prediction and modelling of rainfall-runoff during typhoon events using a physically-based and artificial neural network hybrid model. *Hydrol. Sci. J.* **2014**, *60*, 2102–2116. [CrossRef]
8. Filali Boubrahimi, S.; Neema, A.; Nassar, A.; Hosseinzadeh, P.; Hamdi, S.M. Spatiotemporal data augmentation of MODIS-Landsat water bodies using adversarial networks. *Water Resour. Res.* **2024**, *60*, e2023WR036342. [CrossRef]
9. Morovati, R.; Kisi, O. Utilizing hybrid machine learning techniques and gridded precipitation data for advanced discharge simulation in under-monitored river basins. *Hydrology* **2024**, *11*, 48. [CrossRef]

10. Jakeman, A.J.; Littlewood, I.G.; Whitehead, P.G. Computation of the instantaneous unit hydrograph and identifiable component flows with application to two small upland catchments. *J. Hydrol.* **1990**, *117*, 275–300. [CrossRef]

11. Croke, B.F.W.; Jakeman, A.J. A catchment moisture deficit module for the IHACRES rainfall-runoff model. *Environ. Model. Softw.* **2004**, *19*, 1–5. [CrossRef]

12. Jaiswal, R.K.; Ali, S.; Bharti, B. Comparative evaluation of conceptual and physical rainfall–runoff models. *Appl. Water Sci.* **2020**, *10*, 48. [CrossRef]

13. Yoon, H.; Jun, S.-C.; Hyun, Y.; Bae, G.-O.; Lee, K.-K. A comparative study of artificial neural networks and support vector machines for predicting groundwater levels in a coastal aquifer. *J. Hydrol.* **2011**, *396*, 128–138. [CrossRef]

14. Cho, K.; Van Merriënboer, B.; Gulcehre, C.; Bahdanau, D.; Bougares, F.; Schwenk, H.; Bengio, Y. Learning phrase representations using RNN encoder-decoder for statistical machine translation. *arXiv* **2014**, arXiv:1406.1078.

15. Hochreiter, S.; Schmidhuber, J. Long short-term memory. *Neural Comput.* **1997**, *9*, 1735–1780. [CrossRef] [PubMed]

16. Freer, J.; Beven, K.; Ambroise, B. Bayesian estimation of uncertainty in runoff prediction and the value of data: An application of the GLUE approach. *Water Resour. Res.* **1996**, *32*, 2161–2173. [CrossRef]

17. Farokhi, S.; Yaramal, A.; Huang, J.; Khan, M.F.A.; Qi, X.; Karimi, H. Enhancing the performance of automated grade prediction in MOOC using graph representation learning. In Proceedings of the 2023 IEEE 10th International Conference on Data Science and Advanced Analytics (DSAA), Thessaloniki, Greece, 9–13 October 2023; pp. 1–10. [CrossRef]

18. Gao, S.; Huang, Y.; Zhang, S.; Han, J.; Wang, G.; Zhang, M.; Lin, Q. Short-term runoff prediction with GRU and LSTM networks without requiring time step optimization during sample generation. *J. Hydrol.* **2020**, *589*, 125188. [CrossRef]

19. Govindaraju, R.S. Artificial neural networks in hydrology II: Hydrogeologic applications. *J. Hydrol. Eng.* **2000**, *5*, 124–137.

20. Nagar, U.P.; Patel, H.M. Development of Short-Term Reservoir Level Forecasting Models: A Case Study of Ajwa-Pratappura Reservoir System of Vishwamitri River Basin of Central Gujarat. In *Hydrology and Hydrologic Modelling*; Timbadiya, P.V., Patel, P.L., Singh, V.P., Sharma, P.J., Eds.; Springer: Singapore, 2023; Volume 312. [CrossRef]

21. Hosseinzadeh, P.; Nassar, A.; Boubrahimi, S.F.; Hamdi, S.M. ML-Based Streamflow Prediction in the Upper Colorado River Basin Using Climate Variables Time Series Data. *Hydrology* **2023**, *10*, 29. [CrossRef]

22. Shortridge, J.E.; Guikema, S.D.; Zaitchik, B.F. Machine learning methods for empirical streamflow simulation: A comparison of model accuracy, interpretability, and uncertainty in seasonal watersheds. *Hydrol. Earth Syst. Sci.* **2016**, *20*, 2611–2628. [CrossRef]

23. Yaghoubi, B.; Hosseini, S.A.; Nazif, S. Monthly prediction of streamflow using data-driven models. *J. Earth Syst. Sci.* **2019**, *128*, 1–15. [CrossRef]

24. Rosenblatt, F. The perceptron: A probabilistic model for information storage and organization in the brain. *Psychol. Rev.* **1958**, *65*, 386. [CrossRef] [PubMed]

25. McCulloch, W.S.; Pitts, W. A logical calculus of the ideas immanent in nervous activity. *Bull. Math. Biophys.* **1943**, *5*, 115–133. [CrossRef]

26. Sabzi, H.Z.; King, J.P.; Dilekli, N.; Shoghli, B.; Abudu, S. Developing an ANN Based Streamflow Forecast Model Utilizing Data-Mining Techniques to Improve Reservoir Streamflow Prediction Accuracy: A Case Study. *Civ. Eng. J.* **2018**, *4*, 1135. [CrossRef]

27. Vapnik, V.N.; Chervonenkis, A.Y. On the uniform convergence of relative frequencies of events to their probabilities. In *Measures of Complexity: Festschrift for Alexey Chervonenkis*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 11–30.

28. Asefa, T.; Kemblowski, M.; McKee, M.; Khalil, A. Multi-time scale stream flow predictions: The support vector machines approach. *J. Hydrol.* **2006**, *318*, 7–16. [CrossRef]

29. Jibril, M.M.; Bello, A.; Aminu, I.I.; Ibrahim, A.S.; Bashir, A.; Malami, S.I.; Habibu, M.A.; Magaji, M.M. An overview of streamflow prediction using random forest algorithm. *GSC Adv. Res. Rev.* **2022**, *13*, 050–057. [CrossRef]

30. Muhammad, A.U.; Li, X.; Feng, J. Using LSTM GRU and Hybrid Models for Streamflow Forecasting. In *Machine Learning and Intelligent Communications*; Zhai, X., Chen, B., Zhu, K., Eds.; Springer: Cham, Switzerland, 2019; Volume 294, pp. 510–524. [CrossRef]

31. Box, G.E.P.; Jenkins, G.M.; Reinsel, G.C.; Ljung, G.M. *Time Series Analysis: Forecasting and Control*; John Wiley & Sons: Hoboken, NJ, USA, 2015.

32. Alonso Brito, G.R.; Rivero Villaverde, A.; Lau Quan, A.; Ruíz Pérez, M.E. Comparison between SARIMA and Holt-Winters models for forecasting monthly streamflow in the western region of Cuba. *SN Appl. Sci.* **2021**, *3*, 671. [CrossRef]

33. Adnan, R.M.; Yuan, X.; Kisi, O.; Curtef, V. Application of time series models for streamflow forecasting. *Civ. Environ. Res.* **2017**, *9*, 56–63.

34. Bahri, O.; Li, P.; Boubrahimi, S.F.; Hamdi, S.M. Temporal rule-based counterfactual explanations for multivariate time series. In Proceedings of the 21st IEEE International Conference on Machine Learning and Applications (ICMLA), Nassau, Bahamas, 12–14 December 2022; pp. 1244–1249.

35. Bahri, O.; Boubrahimi, S.F.; Hamdi, S.M. Shapelet-based counterfactual explanations for multivariate time series. *arXiv* **2022**, arXiv:2208.10462.

36. Li, P.; Boubrahimi, S.F.; Hamdi, S.M. Motif-guided time series counterfactual explanations. In Proceedings of the International Conference on Pattern Recognition, Montreal, QC, Canada, 21–25 August 2022; pp. 203–215.

37. Boubrahimi, S.F.; Hamdi, S.M.; Ma, R.; Angryk, R. On the mining of the minimal set of time series data shapelets. In Proceedings of the IEEE International Conference on Big Data (Big Data), Virtually, 10–13 December 2020; pp. 493–502.

38. Alshammari, K.; Hamdi, S.M.; Boubrahimi, S.F. Identifying Flare-indicative Photospheric Magnetic Field Parameters from Multivariate Time-series Data of Solar Active Regions. *Astrophys. J. Suppl. Ser.* **2024**, *271*, 39. [CrossRef]

39. Hosseinzadeh, P.; Boubrahimi, S.F.; Hamdi, S.M. Improving Solar Energetic Particle Event Prediction through Multivariate Time Series Data Augmentation. *Astrophys. J. Suppl. Ser.* **2024**, *270*, 31. [CrossRef]

40. Johnson, R.; Filali Boubrahimi, S.; Bahri, O.; Hamdi, S.M. Combining Empirical and Physics-Based Models for Solar Wind Prediction. *Universe* **2024**, *10*, 191. [CrossRef]

41. Li, P.; Bahri, O.; Boubrahimi, S.F.; Hamdi, S.M. SG-CF: Shapelet-guided counterfactual explanation for time series classification. In Proceedings of the 2022 IEEE International Conference on Big Data (Big Data), Osaka, Japan, 17–20 December 2022; pp. 1564–1569.

42. Xu, D.-M.; Hu, X.-X.; Wang, W.-C.; Chau, K.-W.; Zang, H.-F.; Wang, J. A new hybrid model for monthly runoff prediction using ELMAN neural network based on decomposition-integration structure with local error correction method. *Expert Syst. Appl.* **2024**, *238*, 121719. [CrossRef]

43. Xu, T.; Longyang, Q.; Tyson, C.; Zeng, R.; Neilson, B.T. Hybrid Physically Based and Deep Learning Modeling of a Snow Dominated, Mountainous, Karst Watershed. *Water Resour. Res.* **2022**, *58*, e2021WR030993. [CrossRef]

44. Zhao, S.; Fu, R.; Zhuang, Y.; Wang, G. Long-Lead Seasonal Prediction of Streamflow over the Upper Colorado River Basin: The Role of the Pacific Sea Surface Temperature and Beyond. *J. Clim.* **2021**, *34*, 6855–6873. [CrossRef]

45. Vano, J.A.; Udall, B.; Cayan, D.R.; Overpeck, J.T.; Brekke, L.D.; Das, T.; Hartmann, H.C.; Hidalgo, H.G.; Hoerling, M.; McCabe, G.J.; et al. Understanding Uncertainties in Future Colorado River Streamflow. *Bull. Am. Meteorol. Soc.* **2014**, *95*, 59–78. [CrossRef]

46. Ficklin, D.L.; Stewart, I.T.; Maurer, E.P. Climate change impacts on streamflow and subbasin-scale hydrology in the Upper Colorado River Basin. *PLoS ONE* **2013**, *8*, e71297. [CrossRef] [PubMed]

47. Dawadi, S.; Ahmad, S. Changing climatic conditions in the Colorado River Basin: Implications for water resources management. *J. Hydrol.* **2012**, *430*, 127–141. [CrossRef]

48. Smith, J.A.; Baeck, M.L.; Yang, L.; Signell, J.; Morin, E.; Goodrich, D.C. The Paroxysmal Precipitation of the Desert: Flash Floods in the Southwestern United States. *Water Resour. Res.* **2019**, *55*, 9567–9586. [CrossRef]

49. Christensen, N.S.; Wood, A.W.; Voisin, N.; Lettenmaier, D.P.; Palmer, R.N. The Effects of Climate Change on the Hydrology and Water Resources of the Colorado River Basin. *Clim. Change* **2004**, *62*, 337–363. [CrossRef]

50. U.S. Bureau of Reclamation. *Colorado River Basin Water Supply and Demand Study, Executive Summary*; U.S. Department of the Interior: Washington, DC, USA, 2012; p.34. Available online: https://www.usbr.gov/lc/region/programs/crbstudy/finalreport/Executive%20Summary/CRBS_Executive_Summary_FINAL.pdf (accessed on 19 September 2023).

51. U.S. Bureau of Reclamation. *Colorado River Basin Technical Report*; U.S. Department of the Interior: Washington, DC, USA, 2007. Available online: https://www.usbr.gov/ColoradoRiverBasin/ (accessed on 19 September 2023).

52. Xiao, M.; Udall, B.; Lettenmaier, D.P. On the Causes of Declining Colorado River Streamflows. *Water Resour. Res.* **2018**, *54*, 6739–6756. [CrossRef]

53. Hundley, N., Jr. *Water and the West: The Colorado River Compact and the Politics of Water in the American West*; University of California Press: Berkeley, CA, USA, 1975.

54. Painter, T.H.; Skiles, S.M.; Deems, J.S.; Bryant, A.C.; Landry, C.C. Dust radiative forcing in snow of the Upper Colorado River Basin: 1. A 6 year record of energy balance, radiation, and dust concentrations. *Water Resour. Res.* **2012**, *48*. [CrossRef]

55. Liu, Y.; Peters-Lidard, C.D.; Kumar, S.V.; Arsenault, K.R.; Mocko, D.M. Blending satellite-based snow depth products with in situ observations for streamflow predictions in the Upper Colorado River Basin. *Water Resour. Res.* **2015**, *51*, 1182–1202. [CrossRef]

56. Tillman, F.D.; Gangopadhyay, S.; Pruitt, T. *Trends in Recent Historical and Projected Climate Data for the Colorado River Basin and Potential Effects on Groundwater Availability*; Scientific Investigations Report No. 2020-5107; U.S. Geological Survey: Reston, VA, USA, 2020. [CrossRef]

57. Udall, B.; Overpeck, J. The twenty-first century Colorado River hot drought and implications for the future. *Water Resour. Res.* **2017**, *53*, 2404–2418. [CrossRef]

58. Dettinger, M.D.; Cayan, D.R.; Meyer, M.K.; Jeton, A.E. Simulated hydrologic responses to climate variations and change in the Merced, Carson, and American River basins, Sierra Nevada, California, 1900–2099. *Clim. Chang.* **2004**, *62*, 283–317. [CrossRef]

59. Barnett, T.P.; Adam, J.C.; Lettenmaier, D.P. Potential impacts of a warming climate on water availability in snow-dominated regions. *Nature* **2005**, *438*, 303–309. [CrossRef]

60. Stewart, I.T.; Cayan, D.R.; Dettinger, M.D. Changes toward earlier streamflow timing across western North America. *J. Clim.* **2005**, *18*, 1136–1155. [CrossRef]

61. Hidalgo, H.G.; Das, T.; Dettinger, M.D.; Cayan, D.R.; Pierce, D.W.; Barnett, T.P.; Bala, G.; Mirin, A.; Wood, A.W.; Bonfils, C.; et al. Detection and attribution of streamflow timing changes to climate change in the western United States. *J. Clim.* **2009**, *22*, 3838–3855. [CrossRef]

62. Clow, D.W. Changes in the timing of snowmelt and streamflow in Colorado: A response to recent warming. *J. Clim.* **2010**, *23*, 2293–2306. [CrossRef]

63. Franz, K.J.; Hartmann, H.C.; Sorooshian, S.; Bales, R. Verification of National Weather Service Ensemble Streamflow Predictions for Water Supply Forecasting in the Colorado River Basin. *J. Hydrometeor.* **2003**, *4*, 1105–1118. [CrossRef]

64. Pagano, T.C.; Garen, D.C.; Perkins, T.R.; Pasteris, P.A. Daily updating of operational statistical seasonal water supply forecasts for the western US. *JAWRA J. Am. Water Resour. Assoc.* **2009**, *45*, 767–778. [CrossRef]

65. Werner, K.; Yeager, K. Challenges in forecasting the 2011 runoff season in the Colorado Basin. *J. Hydrometeorol.* **2013**, *14*, 1364–1371. [CrossRef]

66. United States Department of Agriculture. Upper Colorado River Basin SNOTEL Snow/Precipitation Update Report. 2023. Available online: https://wcc.sc.egov.usda.gov/reports/UpdateReport.html?textReport=Upper+Colorado+River+Basin& textRptKey=16&textFormat=SNOTEL+Snow%2FPrecipitation+Update+Report&StateList=Select+a+State&RegionList=16 &SpecialList=Select+a+Special+Report&MonthList=March&DayList=7&YearList=2023&FormatList=N0&OutputFormatList= HTML&textMonth=March&textDay=7&CompYearList=select+a+year (accessed on 27 December 2023).

67. Mote, P.W. Trends in snow water equivalent in the Pacific Northwest and their climatic causes. In *Geophysical Research Letters*; Wiley Online Library: Hoboken, NJ, USA, 2003; Volume 30. [CrossRef]

68. Ahmadi, A.; Olyaei, M.; Heydari, Z.; Emami, M.; Zeynolabedin, A.; Ghomlaghi, A.; Daccache, A.; Fogg, G.E.; Sadegh, M. Groundwater level modeling with machine learning: A systematic review and meta-analysis. *Water* **2022**, *14*, 949. [CrossRef]

69. Hikouei, I.S.; Eshleman, K.N.; Saharjo, B.H.; Graham, L.L.B.; Applegate, G.; Cochrane, M.A. Using machine learning algorithms to predict groundwater levels in Indonesian tropical peatlands. *Sci. Total Environ.* **2023**, *857*, 159701. [CrossRef] [PubMed]

70. Tao, H.; Hameed, M.M.; Marhoon, H.A.; Zounemat-Kermani, M.; Heddam, S.; Kim, S.; Sulaiman, S.O.; Tan, M.L.; Sa'adi, Z.; Mehr, A.D.; et al. Groundwater level prediction using machine learning models: A comprehensive review. *Neurocomputing* **2022**, *489*, 271–308. [CrossRef]

71. Thapa, S.; Zhao, Z.; Li, B.; Lu, L.; Fu, D.; Shi, X.; Tang, B.; Qi, H. Snowmelt-driven streamflow prediction using machine learning techniques (LSTM, NARX, GPR, and SVR). *Water* **2020**, *12*, 1734. [CrossRef]

72. Hu, Y.; Yan, L.; Hang, T.; Feng, J. Stream-flow forecasting of small rivers based on LSTM. *arXiv* **2020**, arXiv:2001.05681.

73. Le, X.-H.; Nguyen, D.-H.; Jung, S.; Yeon, M.; Lee, G. Comparison of Deep Learning Techniques for River Streamflow Forecasting. *IEEE Access* **2021**, *9*, 71805–71820. [CrossRef]

74. Terzi, Ö.; Küçüksille, E.U.; Baykal, T.; Taylan, E.D. Deep and machine learning for daily streamflow estimation: A focus on LSTM, RFR and XGBoost. *Water Pract. Technol.* **2023**, *18*, 2401–2414. [CrossRef]

75. Al-Mukhtar, M. Random forest, support vector machine, and neural networks to modelling suspended sediment in Tigris River-Baghdad. *Environ. Monit. Assess.* **2019**, *191*, 673. [CrossRef]

76. Sibtain, M.; Bashir, H.; Nawaz, M.; Hameed, S.; Azam, M.I.; Li, X.; Abbas, T.; Saleem, S. A multivariate ultra-short-term wind speed forecasting model by employing multistage signal decomposition approaches and a deep learning network. *Energy Convers. Manag.* **2022**, *263*, 115703. [CrossRef]

77. Yap, Z.N.; Musa, S. Stream Flow Forecasting on Pahang River by Time Series Models, ARMA, ARIMA and SARIMA. *Recent Trends Civ. Eng. Built Environ.* **2023**, *4*, 331–341.

78. Khodakhah, H.; Aghelpour, P.; Hamedi, Z. Comparing linear and non-linear data-driven approaches in monthly river flow prediction, based on the models SARIMA, LSSVM, ANFIS, and GMDH. *Environ. Sci. Pollut. Res.* **2022**, *29*, 21935–21954. [CrossRef] [PubMed]