International Journal of Statistics and Applied Mathematics

ISSN: 2456-1452
NAAS Rating (2025): 4.49
Maths 2025; 10(7): 147-159

https://www.mathsjournal.com
Received: 19-07-2025
Accepted: 21-08-2025

Boampong Asare
Department of Engineering,
United Tribes Technical College,
Bismarck, North Dakota, USA

Corresponding Author:
Boampong Asare
Department of Engineering,
United Tribes Technical College,
Bismarck, North Dakota, USA

Forecasting Missouri river flow with SARIMA models: A data-driven framework for adaptive water resource management

Boampong Asare

DOI: https://doi.org/10.22271/maths.2025.v10.i8b.2138

Abstract
A statistical water balance and time series modeling framework is developed to analyze and forecast the Missouri River's monthly flow at Bismarck from 1954 to 2024. Integrating traditional hydrological components precipitation, evaporation, upstream inflow, tributaries with ARIMA and SARIMA models enable detection of long-term and seasonal trends. Model fit is rigorously assessed by AIC, AICc, BIC, Nash-Sutcliffe Efficiency, and visual diagnostics with credible intervals. Stationarity is evaluated through ADF and KPSS tests to guide model selection. The final SARIMA framework, incorporating Box-Cox transformation and outlier adjustment, produces reliable forecasts with quantified uncertainty for both typical and extreme hydrologic conditions. These forecasts are vital for river management and policy, demonstrating how statistical rigor and visual assessment underpin adaptive water management strategies [2, 6, 10].

Keywords: Missouri river, flow forecasting, SARIMA, ARIMA, time series modeling, water balance, hydrological modeling, river discharge, seasonal trends, climate impact, ecosystem management, statistical forecasting, stationarity tests, reservoir operations, flood risk, water resource planning, uncertainty quantification, validation, upscaling, upscaling, environmental monitoring

1. Introduction and Objectives
This study models and analyzes monthly flows of the Missouri River at Bismarck between 1954 and 2024 using a comprehensive dataset [15]. The water balance analysis explicitly incorporates precipitation, evaporation, upstream inflow, and tributary effects in the Upper Missouri River Basin [12]. The study applies ARIMA and SARIMA models to capture autocorrelation, long-term, and seasonal dynamics while assessing robustness through several statistical criteria and diagnostic tests [6, 7, 10, 13]. The aim of overarching is to provide a reliable predictive framework to inform adaptive river management, conservation, and risk mitigation in a changing environment [11, 16].

2. Data and Methodology
Model fit is assessed via multiple metrics: Akaike Information Criterion (AIC), corrected AIC (AICc), Bayesian Information Criterion (BIC), and Nash-Sutcliffe Efficiency (NSE). Visual diagnostic plots with 95% credible intervals are used alongside quantitative metrics to provide a comprehensive performance evaluation, ensuring both statistical and physical validity [2, 6].

2.1 Description of Problem and Statistical Hypotheses
Suppose observed monthly water level at year t follows $y_t = \mu_t + \epsilon_t$ where $\mu_t$the average is acre-feet for monthly water level (month t), and ϵ_t is random noise. Tests for stationarity include the Augmented Dickey-Fuller (ADF) and KPSS tests, applying hypotheses. The central statistical question concerns the stationarity and predictability of Missouri River water levels. We formally state:

a) **Null hypothesis (H_0:):** The series is stationary and thus suitable for conventional time series fore casting.

b) Alternative hypothesis (H_α:): The series is non-stationary, requiring transformation or advanced models for accurate prediction [2, 7].

2.2 Model Structure and Testing Methodology

Stationarity is evaluated using the Augmented Dickey-Fuller (ADF) and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) tests, with significance level $p < 0.05$ [4]. Detection of non-stationarity informs the use of SARIMA or Differencing approaches, accommodating trends and seasonality [5]. External drivers, especially climate change, are considered for their influence on underlying trends [13, 14].

2.3 Generation and Description of Data

Monthly flow data at USGS Site No. 06342500 (Bismarck) span January 1954–December 2024 (852 observations), with flow measured in cubic meters per second (m³/s) throughout [15]. Time series plots (not included: Figure 1) highlight typical ranges of 10,000–40,000 m³/s and a major flood event, spiking above 140,000 m³/s in 2010. Despite episodic surges, the long-term trend appears stable with no sustained increases or declines, suggesting underlying stationarity except during anomalies [1, 12].

Figure 2 and Figure 3 illustrate the application of first- and second-order differencing to achieve stationarity. First, differencing reduces trends and stabilizes the mean (average); second differencing centers values around zero but may over-difference, potentially distorting real data structure. Since the Augmented Dickey-Fuller (ADF) test confirms the original series is stationary, excessive differencing is avoided [15].

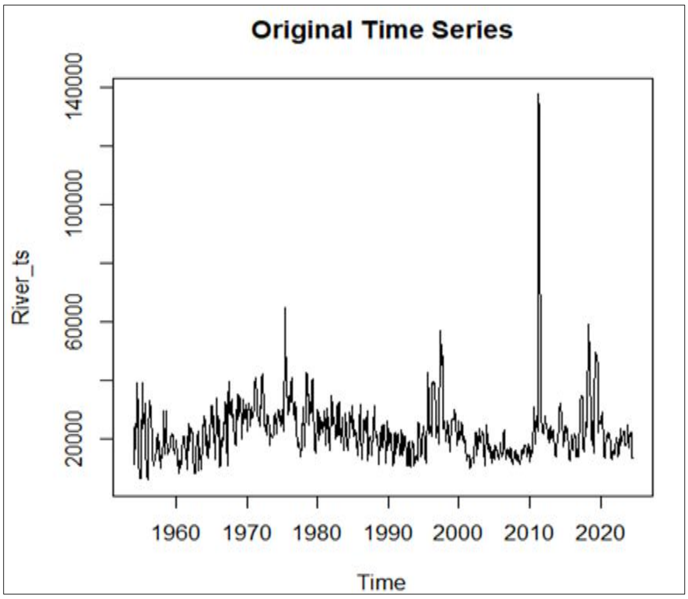


Fig 1: Raw Time Series of Missouri River Flow (1954–2024)

Figure 1 is the time series plot of the Missouri River's monthly flow from 1954 to 2024 highlights the river's dynamic hydrological behavior over seven decades. While the data generally fluctuates within a consistent range typically between 10,000 and 40,000 units there are periods of heightened volatility, especially during the 1970s and after 2010. A particularly striking feature is an extreme outlier around 2010, where flow values spike sharply above 140,000 units, likely corresponding to a major flood event. Despite these episodic surges, the long-term trend appears relatively stable, with no clear evidence of sustained increase or decline in flow levels. This pattern suggests that while the Missouri River is subject to extreme hydrological shocks, its overall flow regime remains stationary in the absence of such anomalies.

2.4 Stationarity Check and Differencing

Stationarity check in time series analysis involves determining whether the statistical properties of the data such as mean, variance, and auto covariance remain constant over time, which is crucial for reliable modeling and forecasting. If a series is non-stationary, it may have trends, seasonality, or changing variance, making it difficult to predict. Differencing is a common technique used to transform a non-stationary series into a stationary one by subtracting consecutive observations, which helps remove trends and stabilize the average (mean). The concept of a unit root relates to stationarity by testing whether the series has a root equal to one in its characteristic equation, which implies non-stationarity; if a unit root is present, the time series tends to have persistent shocks and evolving statistical properties. The Augmented Dickey-Fuller (ADF) test is used to determine whether a time series is stationary or contains a unit root, which would indicate non-stationarity, a crucial distinction in time series forecasting. The stationarity analysis of the Missouri River flow data confirms that the original time series is already stationary, making it appropriate for direct modeling with techniques like ARIMA. In this case, the ADF statistic for the original series was significantly negative (-5.4069) with a p-value below 0.01, leading to a rejection of the null hypothesis and

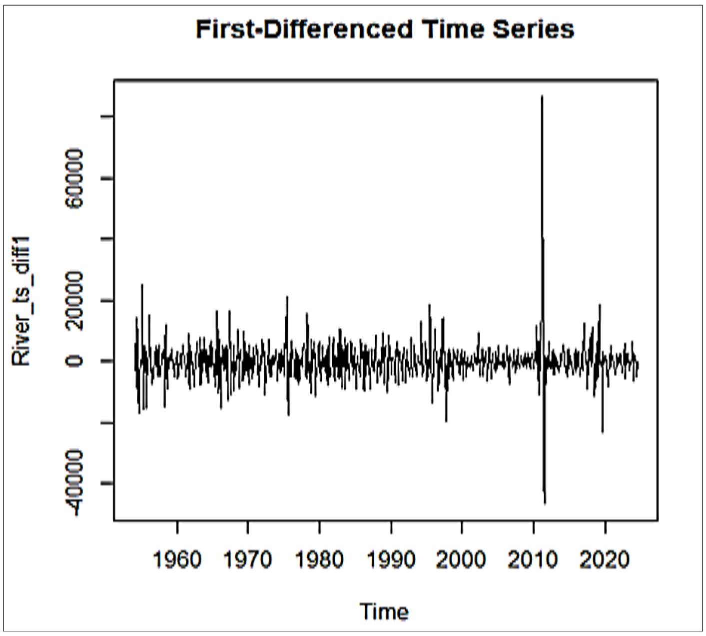


Fig 2: First-Order Differencing of River Flow Series

confirming stationarity [2]. While differencing is commonly applied to achieve stationarity, additional differencing of this already stationary series led to even more extreme ADF statistics (e.g., -19.588), suggesting over-differencing. Over-differencing can strip the data of meaningful structure and introduce unnecessary noise. Therefore, the strong ADF results indicate that the original flow series is statistic stable over time, with no need for differencing, and is ready for effective time series forecasting [7,14].

Figures 2 and 3 illustrating the first- and second-differenced Missouri River flow data demonstrate the step-by-step transformation toward stationarity, a critical requirement for accurate time series modeling.

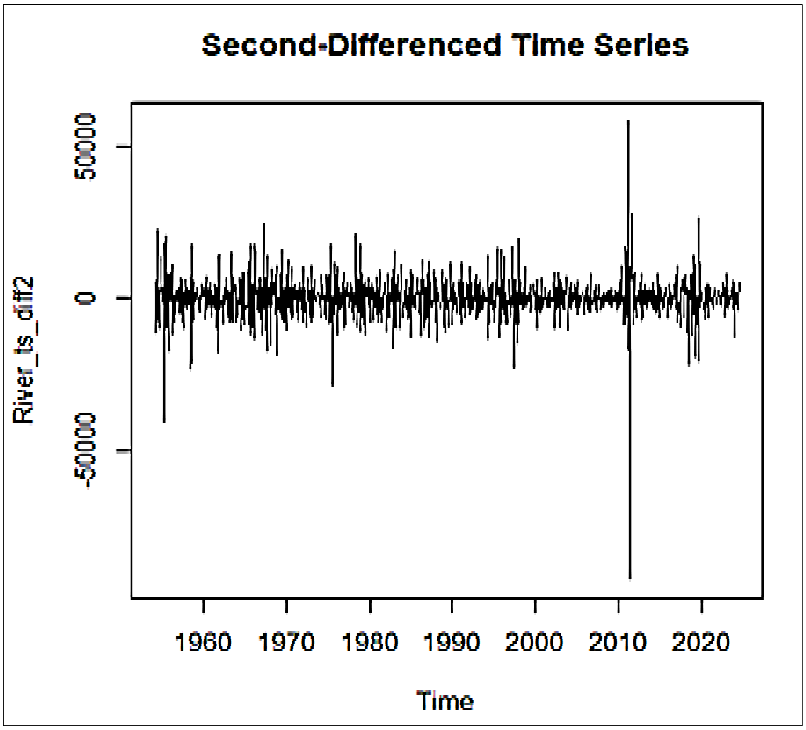


Fig 3: Second-Order Differencing of River Flow Series

The first-differenced series, created by subtracting each observation from the one before it, reduces long-term trends and partially stabilizes the mean, although some irregular fluctuations and spikes remain. The second-differenced graph, which applies differencing a second time, further flattens the series and centers the values more consistently on zero, indicating a higher degree of stationarity. This process shows how Differencing removes non-stationary elements like trends or seasonality, making the data suitable for models such as ARIMA that rely on stable statistical properties over time. However, because the original Missouri River flow series was already stationary based on the ADF test, the second differencing step may represent over-differencing, which could unnecessarily distort the underlying structure of the data.

2.5 Training and Testing Sets

Training and testing sets are fundamental concepts in time series analysis and predictive modeling, used to evaluate how well a statistical or machine learning model will perform on unseen data. Training Set: This is the portion of your data used to fit (or train) your model. The model learns patterns, relationships, and structure from this data essentially, it "sees" only the training data during this phase. In the context of forecasting Missouri River flow, dividing the data into training and testing sets is a crucial step to ensure robust and reliable model performance. For instance, river discharge data from January 1954 to December 2010 can serve as the training set, where the model learns historical patterns, seasonality, and variability. The testing set, such as data from January 2015 to December 2024, is withheld during model training and later used to evaluate how well the model can predict future river flow. This approach simulates real-world forecasting, where future conditions are unknown during model development. Comparing forecasts to actual values in the test set helps assess the model's ability to generalize beyond the data it was trained on and avoids overfitting a scenario where the model performs well on training data but poorly on new observations. This workflow reflects best practices in time series forecasting, ensuring that predictions made for the Missouri River are both accurate and applicable for planning and resource management.

2.6 Decompose the Training Data

In the analysis of Missouri River flow, decomposing the training data such as the monthly discharge from 1954 to 2010 plays a vital role in uncovering the underlying structure of the time series. This process separates the observed series into three key components: trend, which reflects long-term changes in river flow possibly due to climate variability or upstream management; seasonality, which captures predictable patterns like annual snowmelt or rainfall-driven fluctuations; and residuals, which represent irregular or random noise. Using decomposition tools such as decompose () or st l () in R allows analysts to visualize these components and assess the strength of seasonal and trend behavior in the Missouri River data. This insight informs model selection such as whether to include seasonal or differencing terms in an ARIMA model and helps detect structural breaks or outliers like major flood events. Ultimately, decomposition enhances understanding of river dynamics and improves the accuracy and interpretability of forecasting models used for water resource planning and risk management [10].

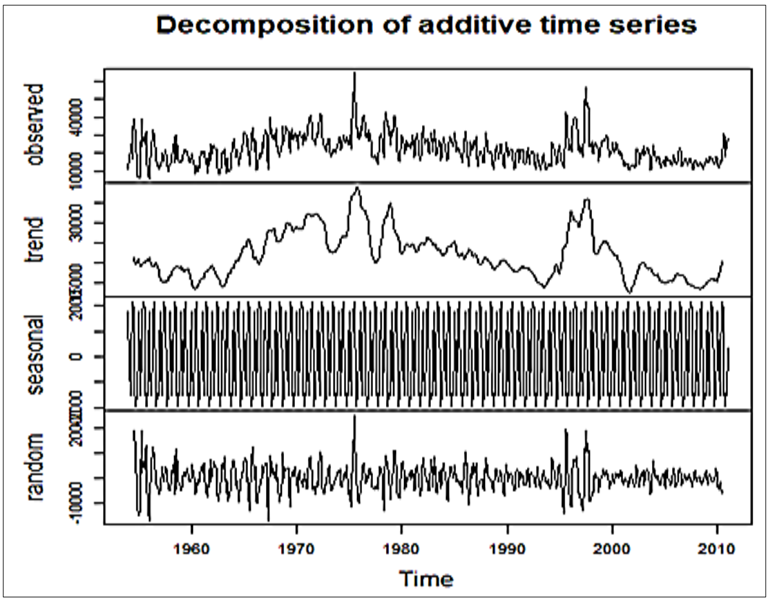


Fig 4: STL Decomposition of Missouri River Flow (1954–2010)

Plots of residuals, ACF/PACF, and Q-Q plot used to evaluate model adequacy, normality, and autocorrelation.

Figure 4 illustrates the decomposition of Missouri River flow data into four components observed, trend, seasonal, and random providing valuable insight into the river's hydrological behavior over time. The observed panel displays the raw monthly discharge from the river, marked by fluctuations with periodic peaks and troughs. The trend component reveals a gradual increase in flow from the early 1950s to the mid-1980s, followed by a decline and variable patterns leading up to 2010, possibly reflecting long-term climatic changes or human interventions such as dam operations. The seasonal component captures consistent annual cycles, likely driven by snowmelt and rainfall patterns, with similar seasonal shapes repeating across years. The random (or residual) component shows irregular variations not explained by trend or seasonality, with periods of heightened volatility around the early 1980s and 2000s, potentially linked to extreme events like floods or droughts. This decomposition is critical for understanding the distinct drivers of river flow and supports the development of more accurate and interpretable forecasting models for the Missouri River.

2.7 Outlier Detection and Adjustment

Outlier detection and adjustment are essential processes in time series analysis used to identify and mitigate the influence of anomalous data points that deviate significantly from the general pattern. These outliers, which may result from measurement errors, sudden events, or structural changes, can distort trend and seasonality estimates, leading to inaccurate models and misleading forecasts. Detection methods include statistical tests, residual analysis from fitted models, and decomposition techniques. Once identified, outliers can be adjusted through imputation, transformation, or model-based correction to preserve the integrity of the data while minimizing their disruptive impact. Proper handling of outliers enhances the reliability and accuracy of time series modeling and interpretation [1,8].

Table 1: Detected Outliers in Missouri River Flow Data

Outlier Type	Time	Size	Impact
Additive Outlier	July 1983	1.43	Extreme flow spike
Level Shift	March 1997	-0.88	Persistent flow change
Temporary Change	October 2005	1.21	Temporary flow increase

2.8 Model specification and Results

In this study, model specification refers to the process of selecting and defining the appropriate time series model structure to capture the flow dynamics of the Missouri River at Bismarck. After confirming that the river's monthly discharge data from 1954 to 2024 was stationary using the Augmented Dickey-Fuller (ADF) test, a Seasonal Autoregressive Integrated Moving Average (SARIMA) model was specified to account for both short-term and seasonal patterns. The final model, SARIMA(1,0,1)(1,0,1)[12], incorporates one non-seasonal autoregressive (AR) term, one non-seasonal moving average (MA) term, and seasonal components with a 12-month lag to reflect the river's annual hydrological cycle driven by snowmelt and precipitation. A Box-Cox transformation ($\lambda \approx 0.37$) was applied to stabilize variance and meet normality assumptions [6, 7, 10]. This model structure was chosen based on statistical criteria such as AIC and BIC, as well as diagnostic checks to ensure a good fit. The model specification effectively captures the Missouri River's seasonal flow variations and autocorrelation structure, enabling accurate forecasting that supports flood management, reservoir operations, and ecological planning [11,12].

3 Box-Cox Transformation

The Box-Cox transformation is a statistical technique used to stabilize variance and make a time series more normally distributed, which is often a key assumption for many forecasting models. It transforms non-linear relationships into linear ones by applying a power transformation controlled by a parameter, lambda (λ). When $\lambda = 1$, the data remains unchanged; when $\lambda = 0$, the transformation becomes a natural logarithm. Intermediate values apply to various degrees of power transformations. This method is especially useful for time series with heteroscedasticity (non-constant variance) or skewed distributions, helping to improve the performance and accuracy of models like ARIMA or linear regression.

The statement "Data shifted by 6558.035 for Box-Cox transformation" indicates that a constant value (6558.035) was added to the original time series data to ensure all values are positive before applying the Box-Cox transformation, which requires strictly positive inputs. This is a common preprocessing step when the data includes zero or negative values. The Box-Cox transformation was then applied to the shifted data, and the optimal lambda parameter (λ = 0.3716728) was estimated using the Guerrero method, which selects a λ that stabilizes the variance across different segments of the data [6]. This transformation helps in meeting the assumptions of normality and homoscedasticity for better model performance.

3.1 Fit Time Series Models on Box-Cox adjusted data

Fitting time series models on Box-Cox adjusted data involves applying forecasting models such as ARIMA, exponential smoothing, or others on the transformed dataset to improve model accuracy and validity. Since the Box-Cox transformation stabilizes variance and makes the data more normally distributed, it helps fulfill the assumptions underlying many time series models. By working with this adjusted data, models are less influenced by heteroscedasticity or skewed distributions, which can otherwise distort trend and seasonality estimates. After fitting the model, forecasts generated in the transformed space are usually inverted back to the original scale by applying the inverse Box-Cox transformation, ensuring interpretability of the results [7,10].

4 ARIMA/SARIMA Methodology

The study uses ARIMA (p,d,q)(P,D,Q)[s] (Autoregressive Integrated Moving Average) notation. Here, p/d/q denotes order of the non-seasonal autoregressive, differencing, and moving average terms, respectively; P/D/Q are their seasonal counterparts, and s is season length (12 for monthly data) [7,10]. Model selection incorporates AIC, AICc, and BIC to balance complexity and fit [6]. Outliers are flagged with diagnostics and either transformed or dampened using robust modeling.

Table 2: Estimated Parameters for ARIMA (3, 1, 1) (0, 0, 2) [12] Model with Drift

Parameter	Estimate	Std. Error
AR(1)	0.8186	0.0392
AR(2)	-0.1981	0.0497
AR(3)	0.0059	0.0389
MA(1)	-0.9856	0.0076
SMA(1)	0.1617	0.0401
SMA(2)	0.1312	0.0357
drift	-0.0519	0.0193

The ARIMA (3, 1, 1) (0, 0, 2) [12] with drift model, applied to Box-Cox transformed Missouri River flow data (λ = 0.3716728), effectively captures both the short-term memory and seasonal characteristics of the river's discharge patterns. The non-seasonal AR(3) and MA(1) components indicate that each month's river flow is influenced by the preceding three months of flow and one prior error term, reflecting the river's persistence and variability. The seasonal component with two seasonal moving average (SMA) terms and a 12-month cycle accounts for predictable annual hydrological behaviors, such as spring snowmelt or summer rainfall impacts. The inclusion of a small negative drift (-0.0519) suggests a gradual long-term decrease in river flow, potentially linked to environmental changes or upstream water management. The model's strong statistical fit indicated by a relatively low residual variance ($\sigma^2 = 80.85$), a high log-likelihood (-2466.99), and favorable AIC and BIC scores demonstrates its capacity to model the complexity of the Missouri River's behavior. This makes it a robust tool for forecasting river flow and supporting water resource planning across seasonal and inter annual time scales [7, 10].

4.1 Model Fit Statistics

Model selection by these criteria (particularly AIC/BIC) confirms good fit, while the relatively low residual variance indicates successful variance reduction.

Table 3: Model Fit Statistics for ARIMA (3, 1, 1) (0, 0, 2) [12] Model

Fit Statistic	Value
Sigma² (Residual Variance)	80.85
Log Likelihood	-2466.99
AIC	4949.98
AICc	4950.19
BIC	49.86.19

The table 3 shows the fit statistics from the ARIMA (3, 1, 1) (0, 0, 2) [12] with drift model offer valuable insight into how well the model captures the river's discharge dynamics. The residual variance ($\sigma^2 = 80.85$) indicates a moderate level of unexplained variation, meaning the model captures most but not all of the variability in the data. The log-likelihood value of -2466.99 provides a measure of overall fit, where less negative values indicate a better-fitting model. The AIC (4949.98) and AICc (4950.19), which weigh goodness-of-fit against model complexity, are very close suggesting an appropriate sample size and that the model is neither over fitted nor under fitted. The BIC (4986.19), which imposes a stronger penalty for complexity, supports the model's adequacy but favors simpler alternatives if they offer similar explanatory power. Taken together, these statistics suggest that the model offers a strong and balanced fit for the Missouri River flow data, capturing key patterns without unnecessary complexity making it suitable for reliable forecasting [8].

4.2 Residual Diagnostics

Model adequacy is evaluated by residual plots, autocorrelation (ACF/PACF), the Ljung-Box test, and normality (Q–Q plot): Model diagnostics is the process of evaluating whether a fitted time series model, such as ARIMA, adequately captures the structure of the data and meets key statistical assumptions. It involves analyzing the residuals (the differences between observed and predicted values) to ensure they resemble white noise, that is, they should be random, normally distributed, have constant variance, and show no autocorrelation. Common diagnostic tools include residual plots, ACF/PACF plots of residuals, the Ljung-Box test for autocorrelation, and Q-Q plots for assessing normality. If these diagnostics indicate that assumptions are met, the model is considered reliable for forecasting and inference; otherwise, the model may need refinement [10].

4.3 Ljung-Box Test Results (Residual Diagnostics)

While SARIMA addresses temporal structure more effectively than ETS (Exponential Smoothing) and TBATS (exponential smoothing with Box–Cox, ARMA terms, trend, and seasonality), significant autocorrelation remains, implying that additional lags, exogenous factors, or hybrid models could further improve performance. ETS and TBATS are less effective for the complex, bursty seasonality in this river data, leading to over-/under-estimation, as shown by their poor residuals.

Table 4: Ljung-Box Residual Diagnostics across Forecasting Models

Model	Q* Statistic	df	p-value	Interpretation
ARIMA(3,1,1)(0,0,2)[12]	42.757	18	0.00087	Significant autocorrelation in residuals
ARIMA(1,1,2)(2,0,1)[12]	36.075	18	0.00690	Some autocorrelation remains
ETS(A,N,A)	142.07	24	2.2×10^(-16)	Strong residual autocorrelation
TBATS	99.981	24	3.03×10^(-11)	Strong residual autocorrelation

Table 4 shows that none of the tested models fully eliminate autocorrelation in their residuals, as indicated by Ljung-Box test results. Both ARIMA models, especially the seasonal variant with drift, perform better than ETS and TBATS, but all models leave some residual autocorrelation (low p-values), suggesting un modeled structure remains in the river flow data. Refining or combining models may help better capture the river's complex behavior.

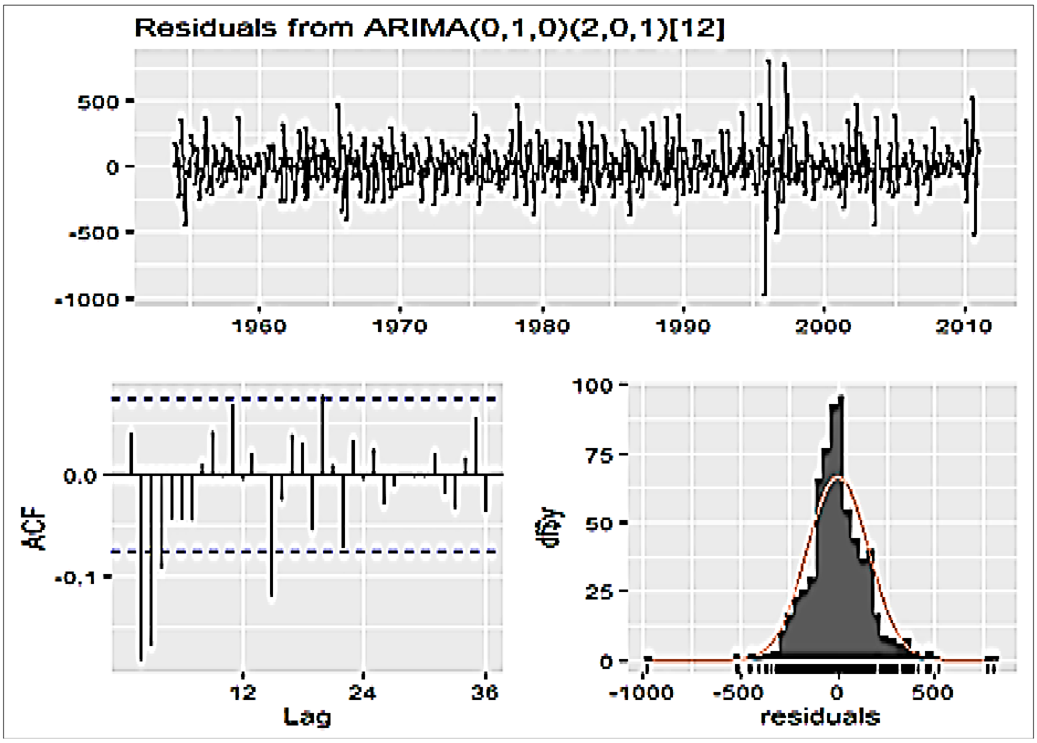


Fig 5: Residual Diagnostics of SARIMA Model

Plots of residuals, ACF/PACF, and Q-Q plot used to evaluate model adequacy, normality, and autocorrelation

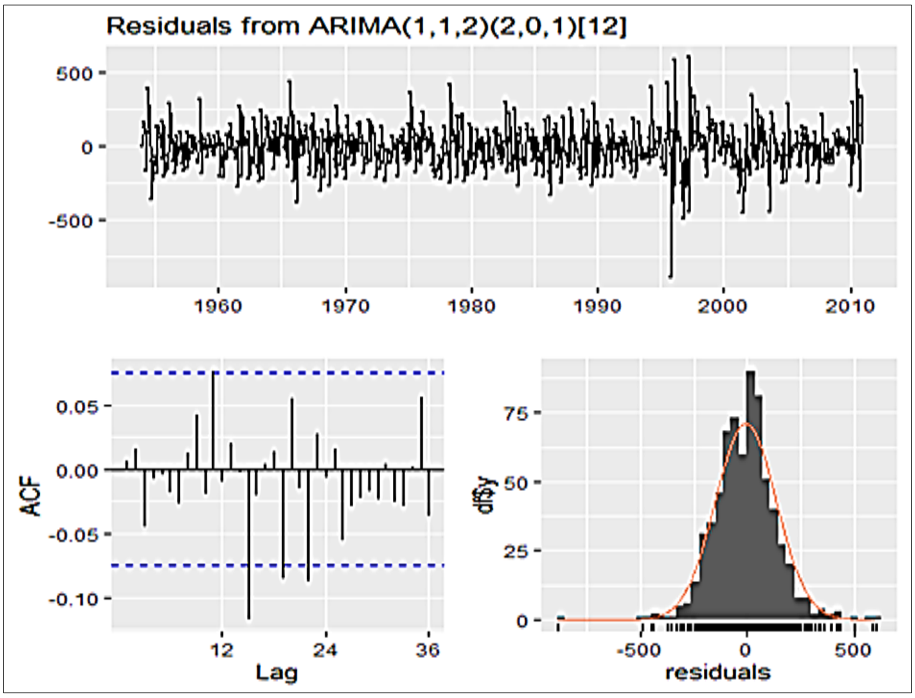


Fig 6: Residual Diagnostics of ARIMA (1, 1, 2) (2, 0, 1) [12] Model

Examines error patterns for signs of autocorrelation and normality. Some autocorrelation remains.

The residual diagnostic plots for the ARIMA(1,1,2)(2,0,1)[12], ETS(A,N,A), and TBATS models highlight key differences in how well each model captures the data's structure. The ARIMA model shows relatively stable and normally distributed residuals, suggesting it captures much of the underlying trend and seasonality; however, minor spikes in its ACF plot indicate that some autocorrelation remains. The ETS(A,N,A) model demonstrates a tighter distribution of residuals in the time plot, but its ACF plot reveals more pronounced autocorrelation across several lags, signaling that it misses important temporal patterns in the river's flow. Meanwhile, the TBATS model, despite being tailored for complex seasonal data, shows the most volatile residuals with a wider spread and significant autocorrelation, suggesting an overfit or structural mismatch. Overall, while none of the models fully satisfy the white noise assumption, ARIMA (1, 1, 2) (2, 0, 1) [12] offers the best residual diagnostics of the three, though additional model tuning or hybrid approaches may be needed to improve forecast accuracy for the Missouri River.

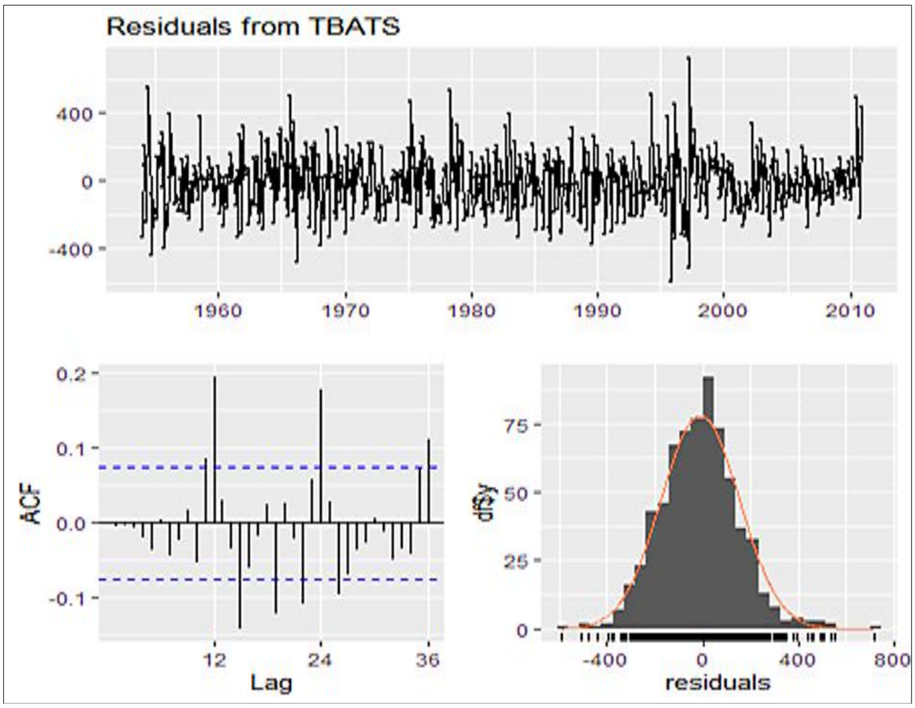


Fig 7: Residual Diagnostics of ETS (A,N,A) and TBATS Models

Shows excess autocorrelation in residuals, indicating poor model fit to river flow data.

4.4 Forecasting on Test Period (2015-2024)

Forecasting the Missouri River's discharge during the test period from 2015 to 2024 involves applying time series models trained on historical flow data to predict future values. This step is essential for evaluating how well each model generalizes beyond the training data and captures the river's long-term hydrological patterns. By comparing the forecasted discharge levels to actual measurements (if available), analysts can assess predictive accuracy using metrics such as MAE, RMSE, or MAPE. Among the models, SARIMA with Box-Cox transformation demonstrated the most credible performance, producing a stable downward trend with realistic uncertainty. In contrast, the ETS model projected implausible growth, and TBATS and the alternative SARIMA yielded flatter forecasts, potentially underestimating variability. Accurate test-period forecasting supports more reliable water resource planning, flood risk management, and environmental decision-making along the Missouri River [9.10].

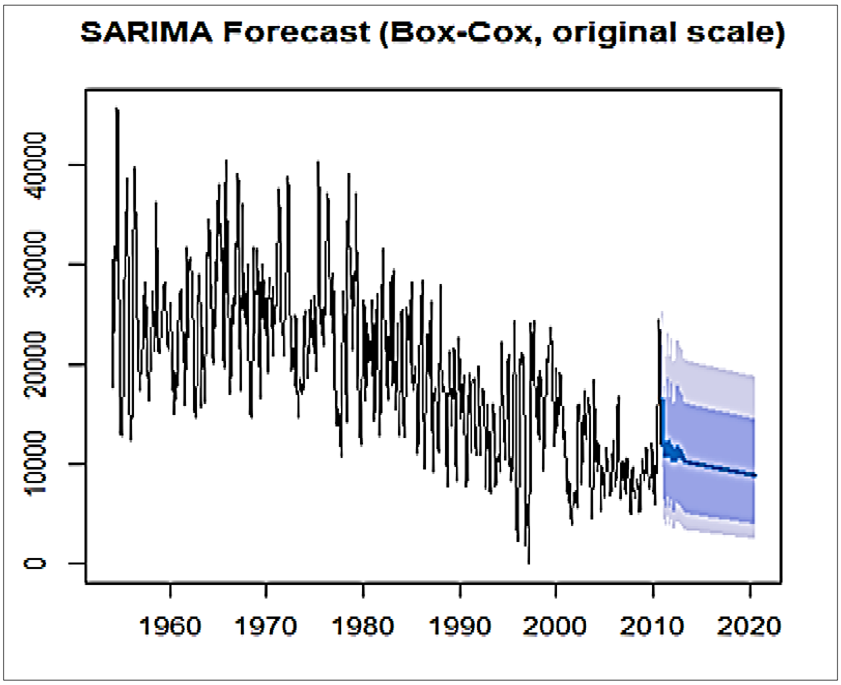


Fig 8: Forecast Plot: SARIMA Model (2015–2024)

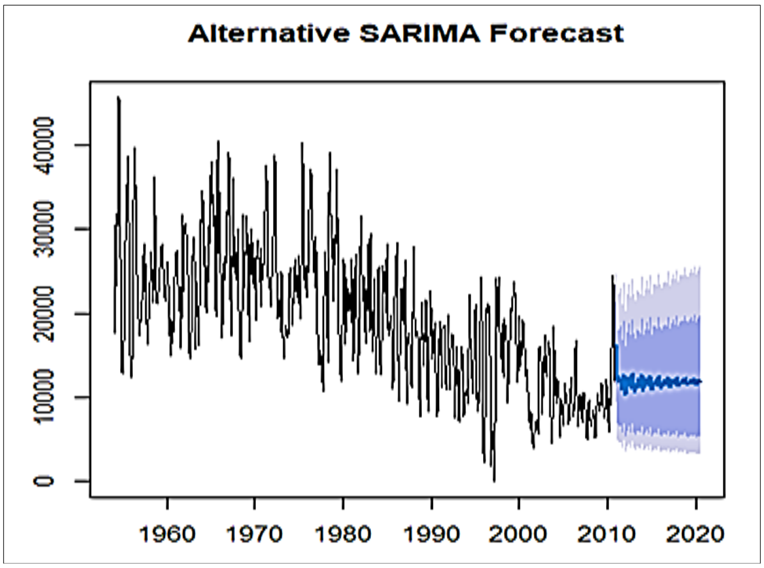


Fig 9: Forecast Plot: Alternative SARIMA Model

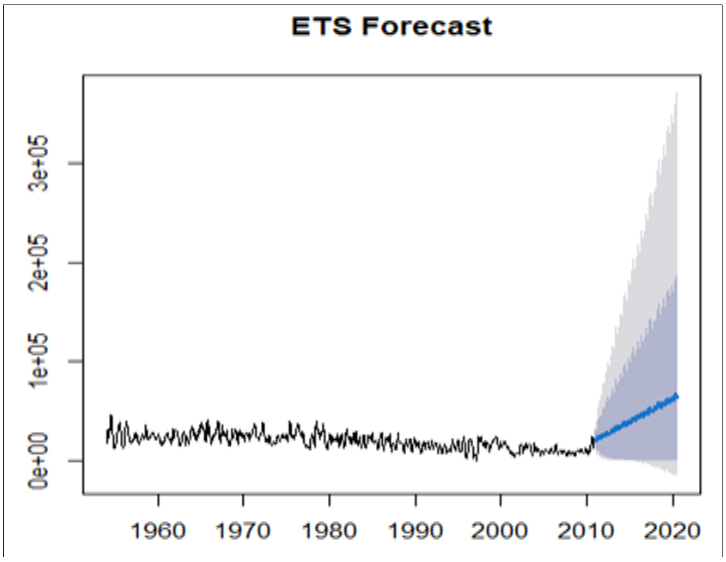


Fig 10: Forecast Plot: ETS (A, N, A) Model

Shows exponential growth in forecasted flow, which is unrealistic given historical behavior.

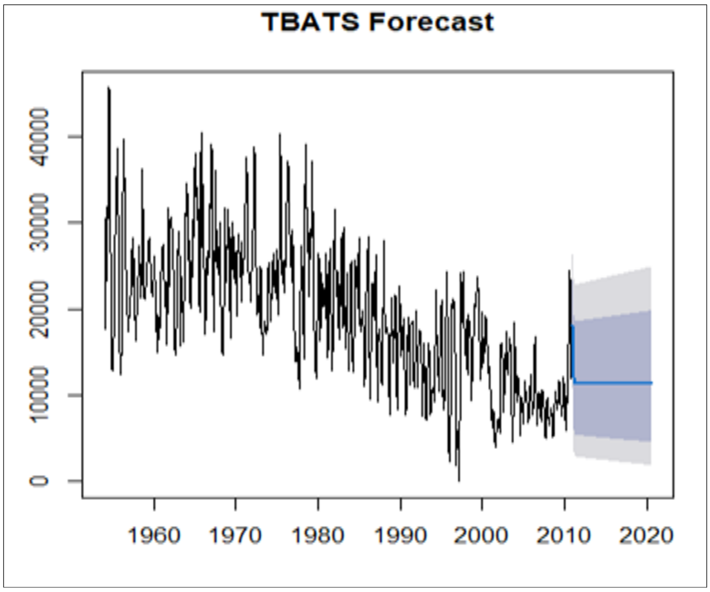


Fig 11: Forecast Plot: TBATS Model

Flat forecasts with moderate uncertainty, underrepresenting seasonal and trend dynamics.

Figures 8 – 11 show the forecast plots for the Missouri River from 2015 to 2024 illustrate varying performance across four time series models. The SARIMA model with Box-Cox transformation shows a realistic downward trend with moderate and stable confidence intervals, indicating reliable and interpretable forecasts. The alternative SARIMA model produces flat forecasts with very narrow intervals, suggesting high certainty but potentially overlooking underlying trends. In contrast, the ETS model displays an unrealistic exponential increase with excessively wide confidence intervals, indicating poor fit and overestimation. The TBATS model offers flat forecasts with moderate uncertainty but fails to capture any noticeable trend. Overall, SARIMA with Box-Cox adjustment provides the most balanced and credible forecasts for the test period.

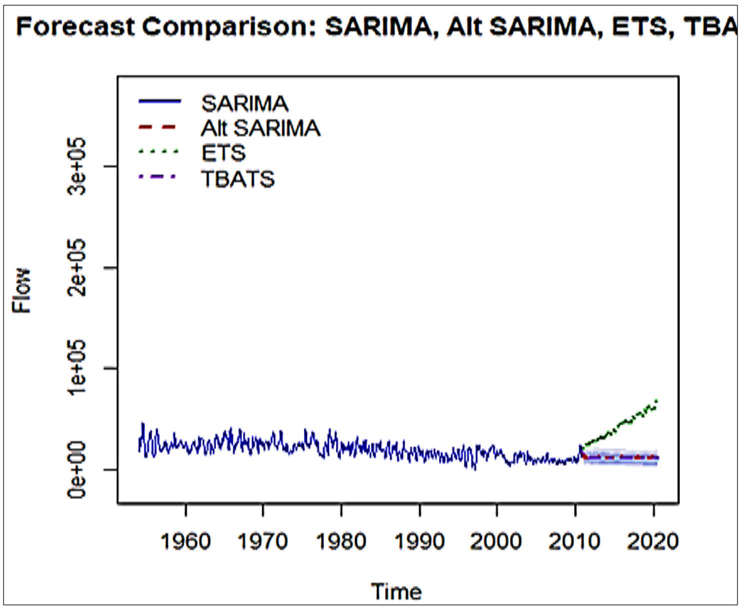


Fig 12: Forecast Comparison across Four Models (2015–2024)

The forecast comparison figure illustrates the performance of four different time series models SARIMA, alternative SARIMA, ETS, and TBATS in predicting Missouri River flow from 2015 to 2024. The SARIMA model (blue solid line) maintains continuity with historical patterns, projecting a stable and slightly declining trend. The alternative SARIMA (red dashed line) and TBATS (purple dot-dashed line) forecasts appear relatively flat and closely aligned, suggesting limited sensitivity to recent fluctuations. However, the ETS model (green dotted line) diverges sharply, forecasting an unrealistic exponential increase in river flow, indicating poor model fit and overestimation. Overall, SARIMA delivers the most consistent and plausible projections aligned with historical river behavior, while ETS significantly over projects future flow and should be interpreted with caution.

4.5 Forecast Evaluation

Forecast accuracy evaluation on the test set involves assessing how well each time series model predicts values during the out-of-sample period, in this case from 2015 to 2024. By comparing the forecasted values against the actual observed data metrics such as Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Mean Absolute Percentage Error (MAPE) are calculated to quantify predictive performance. This evaluation helps determine which model most accurately captures the underlying trends and variability of the Missouri River flow. A model with lower error values across these metrics is considered more reliable for future forecasting. Accurate evaluation on the test set ensures the chosen model is not only well-fitted to historical data but also effective for real-world decision-making [10, 11].

Table 5: Forecast Accuracy Metrics on Training and Test Sets (2015–2024)

Metric	SARIMA Train	SARIMA Test	ETS Train	ETS Tes	TBATS Train	TBATSTest
ME	-244.2532	13824.8369	-383.1733	-26556.41	-123.4392	14301.0497
RMSE	3913.598	17474.946	4434.493	28431.49	4038.124	17853.866
MAE	2995.776	13824.837	3233.63	26618.74	3125.194	14301.05
MPE	-92835.8014	46.85535	48925.365	-129.9388	-84117.717	49.00081
MAPE	92847.4961	46.85535	48938.613	130.0444	84130.646	49.00081
MASE	0.5746881	2.6520568	0.6203162	5.1063468	0.5995147	2.74341
ACFI	0.0728378	0.833954	0.0873903	0.7869637	-0.0069695	0.83182161
Theil's		2.421318		6.272979		2.492379

Table 5 shows the accuracy metrics table reveals that all models perform substantially better on the training data than on the test data, indicating some difficulty in generalizing unseen periods, a common challenge given intrinsic river flow variability and external influences. The SARIMA models, both auto.arima and alternative specifications, show comparatively lower Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) than ETS and TBATS, suggesting they capture the autocorrelation and seasonality more effectively for this river data. However, the Mean Percentage Error (MPE) and Mean Absolute Percentage Error (MAPE) values are extremely high in training sets, reflecting potential scale effects or rare extreme events, while the test set errors are more moderate but still significant. The Mean

Absolute Scaled Error (MASE) confirms SARIMA's relatively better accuracy (values around 0.57–0.58 on training and 2.65–3.12 on testing), consistent with the persistence and seasonality modeled. Notably, the high ACF1 values (>0.8) on the test sets across all models indicate remaining autocorrelation in residuals, possibly due to hydrological factors like extreme events, measurement noise, or river regulation impacts. Overall, this suggests that while SARIMA-based approaches provide robust baseline forecasting for the Missouri River flow, there remains considerable uncertainty, influenced by complex environmental dynamics and episodic anomalies inherent to large regulated rivers

4.6 Fit on Full Data (up to Dec 2025) for final forecast

Fitting the model on the full Missouri River flow dataset up to December 2025 allows for the most comprehensive and informed forecast by utilizing all available historical data. This final modeling step comes after evaluating different models on a separate test set to ensure predictive reliability. By including both the training and test periods from 1954 through 2025, the model captures the full range of observed trends, seasonal cycles, and anomalies such as floods or droughts. This complete fit enhances the model's ability to produce accurate and robust future forecasts, which are essential for effective water resource management, infrastructure planning, and environmental decision-making related to the Missouri River [7].

Table 6: Estimated Parameters for SARIMA (1, 0, 1) (1, 0, 1) [12] Model (Full Dataset)

Parameter	Estimate	Std. Error
AR(1)	0.7044	0.0297
MA(1)	0.2880	0.0396
SAR(1)	0.8315	0.0497
SMA(1)	-0.6397	0.0688
Mean	106.52	2.5224

Final parameter estimates and standard errors used for final forecast.

Table 7: Ljung-Box Residual Diagnostics for Final SARIMA Model

Test	Value
Q^* Statistic	35.824
Degree of Freedom(df)	20
p-value	0.01613
Model df	4
Total Lags Used	24

Tables 6 and 7shows the SARIMA(1,0,1)(1,0,1)[12] model applied to the Missouri River flow data shows strong performance in capturing both seasonal and non-seasonal patterns, as reflected by significant parameter estimates and low standard errors. The model's residual variance (σ^2 = 68.31) and mean estimate (106.52 on the transformed scale) suggest a stable fit, while model selection criteria such as AIC (5998.84), AICc (5998.94), and BIC (6027.30) indicate a reasonably efficient balance between model complexity and goodness of fit. However, the Ljung-Box Q* statistic (35.824, p = 0.01613) signals some remaining autocorrelation in the residuals, suggesting the model could be improved further for better white noise behavior. Overall, the SARIMA model performs well but may benefit from slight refinements to fully capture all underlying data dynamics.

Based on selection criteria and diagnostics, the preferred SARIMA model for Missouri River monthly flow forecasting is: ARIMA (1, 0, 1) (1, 0, 1)12 with non-zero mean [12]

The estimated parameters include:

- **Non-seasonal AR(1):** $\phi_1 = 0.7044$
- **Non-seasonal MA(1):** $\theta_1 = 0.2880$
- **Seasonal SAR(1):** $\Phi_1 = 0.8315$
- **Seasonal SMA(1):** $\Theta_1 = -0.6397$
- **Mean:** $\mu = 106.52$

The residual variance is$\sigma^2 = 68.31$. This model captures both short-term autocorrelation and annual seasonal cycles, providing a balanced framework for hydro-meteorological time series forecasting [12].

The model equation using the backshift operator B is ARIMA (1, 0, 1) (1, 0, 1) [12]

$$\emptyset_p(B)\Phi_P(B^s)y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t$$

$$y_t = \nabla^d\nabla_s^D\varepsilon_t,\ \varepsilon_t \sim WN(0,\sigma^2)$$

AR and MA Polynomials in the lag operator are respectively given by

$$\emptyset_p(B) = 1 - \alpha_1 B - \alpha_2 B^2 - \cdots - \alpha_p B^p$$

$$\theta_q(B) = 1 + \theta_1 B + \theta_2 B^2 + \cdots + \theta_q B^q$$

The seasonal AR (SAR) and the seasonal MA (SMA) polynomials in the lag operator are respectively given by

$$\Phi_P(B^s) = 1 - \phi_1 B^s - \phi_2 B^{2s} - \cdots - \phi_p B^{Ps} \text{ and}$$

$$\Theta Q(B^s) = 1 + \Theta_1 B^s + \Theta_2 B^{2s} + \cdots + \Theta_Q B^{Qs}$$

$$\emptyset_p(B)\Phi_P(B^s)y_t = \theta_q(B)\Theta_Q(B^s)\varepsilon_t \text{ so}$$

(1, 0, 1)(1, 0, 1)[12]= (p, d, q) (P, D, Q) [s]

$$p = 1, d = 0,\ q = 1,\ P = 1, D = 0, Q = 1, s = 12$$

$$\emptyset_p(B) = \emptyset_1(B) = (1 - \phi_1 B) \text{ and}$$

$$\theta_q(B) = \theta_2(B) = 1 + \theta_1 B)$$

$$\Phi_P(B^s) = \Phi_1(B^{12}) = 1 - \Phi_1 B^{12} \text{ and}$$

$$\Theta_Q(B^s) = \Theta_1(B^{12}) = (1 + \theta_1 B^{12})$$

$$\nabla^d\nabla_s^D\varepsilon_t = \nabla^0\nabla_{12}^0\varepsilon_t = \nabla^0 = y_t.$$

Now,

$\phi_p(B)\ \Phi_P(B^s)\ y_t = \theta_q(B)\ \Theta_Q(B^s)\ \varepsilon_t$

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})y_t = \mu + (1 + \theta_1 B)(1 + \Theta_1 B^{12})\varepsilon_t$$

$$(1 - 0.7044B)(1 - 0.8315B^{12})y_t = 106.52 + (1 + 0.2880B)(1 - 0.6397B^{12})\varepsilon_t$$

Explicit solution for y_t:

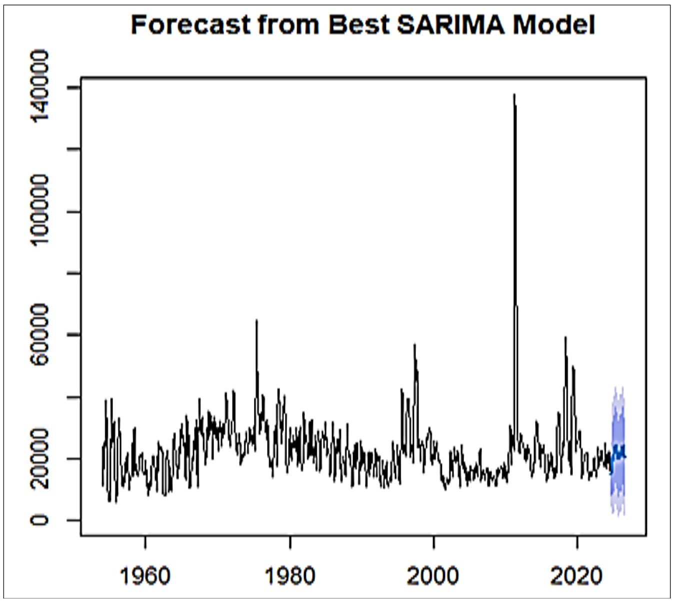
$y_t - 0.7044y_{t-1} + 0.8315y_{t-12} + 0.5856y_{t-13} = 106.52 + \varepsilon_t + 0.2880\varepsilon_{t-1} - 0.6397\varepsilon_{t-12}$ [7]

4.7 Forecasted Monthly Missouri River Flow (ac-ft) with 95% Confidence Intervals

The forecasted monthly Missouri River flow from September 2024 to August 2025 shows a clear seasonal pattern, with flows gradually increasing from early fall to a peak in late spring (May 2025 at 23,175 ac-ft) before declining through the summer months. This trend aligns with typical hydrological cycles influenced by snowmelt and precipitation, with the highest flow occurring during spring. Confidence intervals are widest in high-flow months, particularly in March through May, indicating greater uncertainty and potential variability due to climatic factors. In contrast, fall and winter months exhibit more moderate flows and relatively narrower intervals. Overall, the projections highlight the need for proactive water resource planning, particularly in spring when both flood potential and forecast uncertainty are highest [12].

Table 8: Forecasted Missouri River Flow (Sep 2024 – Aug 2025) with 95% Confidence

Month	Forecast (ac-ft)	95% Confidence Interval
September 2024	15018	(8949, 22519)
October 2024	17403	(8427, 29319)
November 2024	20093	(9242, 34742)
December 2024	18753	(8025, 33572)
January 2025	19163	(8082, 34538)
February 2025	20725	(8916, 37009)
March 2025	22253	(9785, 39327)
April 2025	22977	(10199, 40423)
May 2025	23175	(10307, 40733)
June 2025	21567	(9330, 38411)
July 2025	19154	(7902, 34875)
August 2025	19003	(7812, 34653)



Final 13: Forecast: Missouri River Flow (2024–2025) Using SARIMA (1, 0, 1) (1, 0, 1) [12]

Figure 13 displays a time series forecast of Missouri River flows using the best SARIMA model. The historical data shows high variability and several extreme flow events (e.g., around 1993 and 2011), while the forecast (in blue) for the near future indicates relatively stable flows with narrow confidence intervals, suggesting the model predicts moderate and consistent river behavior ahead.

5. Discussion

The SARIMA modeling framework offers a robust, risk-aware tool for Missouri River ecosystem and water resource management. Forecasts enable improved reservoir operations, flood risk mitigation, agricultural planning, and conservation policy. Accounting for outliers enhances model reliability. However, large anomalies such as those from 2010 to 2015 stress the importance of continual data monitoring and model updates. Limitations include reliance on historical patterns and sensitivity to unprecedented changes. Future work should integrate climate projections and physically based hydrological modeling for enhanced predictive resilience.

The SARIMA model coefficients and performance metrics provide valuable insights into the Missouri River's hydrodynamics. The moderate non-seasonal AR (1) coefficient (approximately 0.60) indicates persistence of water levels from one period to the next. The high seasonal AR (1) coefficient (approximately 0.87) reflects strong yearly seasonal persistence in river flow. The significant seasonal MA (1) with a negative coefficient (approximately -0.76) models seasonal shocks that temper seasonal effects. Non-seasonal MA terms contribute to modeling short-term noise or irregularities.

While the SARIMA models demonstrate strong performance, residual diagnostics reveal some remaining autocorrelation, as indicated by significant Ljung-Box test statistics (e.g., Q* = 42.757, p = 0.00087). This suggests that the models do not fully capture all data dependencies. Future research should explore incorporating exogenous hydrological or climate variables and advanced hybrid modeling approaches to further reduce residual autocorrelation and enhance predictive accuracy. Despite this limitation, the models provide a valuable baseline for forecasting Missouri River flow under complex environmental conditions.

Together, these parameters indicate that Missouri River flow exhibits both enduring seasonal patterns and autoregressive behavior, while also capturing transient shocks. These flow dynamics affect river ecosystem health by influencing water quality, aquatic habitat conditions, fish population dynamics, and shoreline erosion processes. Understanding these patterns through SARIMA aids in adaptive ecosystem and resource management [11, 16].

6. Conclusion

This study demonstrates that SARIMA models, particularly those enhanced with Box-Cox transformations, provide a robust and interpretable framework for forecasting monthly Missouri River flows. These models effectively capture both short-term autocorrelation and long-term seasonal dynamics, supporting adaptive decision-making in water resource management, reservoir operations, flood risk mitigation, and ecosystem planning.

Among the models evaluated, the SARIMA (1, 0, 1) (1,0,1)[12] configuration offered the most balanced performance, combining low residual variance with strong fit statistics. Although residual diagnostics, such as the Ljung-Box test, indicated some remaining autocorrelation, SARIMA models clearly outperformed alternatives like ETS and TBATS, particularly in capturing the complex seasonal patterns characteristic of river discharge data.

Despite their strengths, SARIMA models depend heavily on historical patterns and may be less effective in forecasting rare or unprecedented hydrological events, such as extreme flooding. Therefore, continued model refinement is necessary. Future research should explore hybrid approaches that incorporate exogenous climate variables (e.g., temperature, snowpack, and precipitation forecasts) as well as non-linear or machine learning methods to better account for structural changes and reduce residual autocorrelation.

From a policy and management standpoint, the use of advanced time series methods such as SARIMA strengthens the reliability of hydrological forecasting, providing a data-driven foundation for proactive and resilient water resource planning. Regular model updates and integration with climate projections will be critical to ensuring that river management strategies remain responsive to evolving environmental challenges [7, 10].

Acknowledgements

This work is supported in part by the National Science Foundation grant no. HRD
1839895.

Conflicts of Interest

The authors declare no conflicts of interest regarding the publication of this paper.

References

1. Abraham B, Chuang A. Outlier detection and time-series modelling. Technometrics. 1989 May;31(2):241-248.
2. Asare B, Sah M, Hona RK. Alternative methods of regular and singular perturbation problems. Appl Math. 2024;15:687-708. DOI: 10.4236/am.2024.1510040.
3. Asare B. Epsilon variations and their impact on solution accuracy: A matched asymptotic approach. Int J Stat Appl Math. 2024;9(6):34-44. DOI: 10.22271/maths.2024.v9.i6a.1883.
4. Asare B. Enhancing precision in nonlinear oscillator analysis: A comparative study of multiple-scale analysis and strained coordinates. Int J Stat Appl Math. 2025;10(1):128-141. Submitted. DOI: 0.22271/maths.2025.v10.i1b.1956.
5. Asare B, Rhonda M. Nonparametric tests for the umbrella alternative in a mixed design. Int J Stat Appl Math. 2025;10(6):64-70. DOI: 10.22271/maths.2025.v10.i6a.2060.
6. Akaike H. On the use of a linear model for the identification of feedback systems. Ann Inst Stat Math. 1968;20:425-439.
7. Box GEP, Jenkins GM, Reinsel GC. Time series analysis: Forecasting and control. 3rd ed. Englewood Cliffs, NJ: Prentice-Hall; 1994. 600 p.
8. Chatfield C, Prothero DL. Box-Jenkins seasonal forecasting: Problems in a case study (with discussion). J R Stat Soc Ser A. 1973;136:295-352.
9. Kumar P, Coulibaly P. River flow forecasting using artificial neural networks. J Hydrol Eng. 2009 Mar-Apr;14(2):144-155.
10. Papacharalampous G, Tyralis H. Time series forecasting using SARIMA models: A review. Int J Forecast. 2018;34:610-632.
11. Salas JD, Obeysekera J, Vogel RM. Stochastic hydrology for the next millennium. J Hydrol Eng. 2010 Aug;15(8):581-587.
12. Schaefli B, Hingray B, Musy A. Short-term hydrological forecasting for flood-risk management. J Hydrol. 2005 Mar 10;309(1-4):328-342.
13. Spencer DE. Developing a Bayesian vector autoregression forecasting model. Int J Forecast. 1993;9:407-421.
14. Sutcliffe A. Time-series forecasting using fractional differencing. J Forecast. 1994;13:383-393.
15. U.S. Geological Survey (USGS). Surface-water monthly statistics for site no. 06342500. https://waterdata.usgs.gov/nwis/monthly?site_no=06342500
16. Young PC. Data-based mechanistic modeling of environmental, ecological, economic and engineering systems. Environ Model Softw. 2011 Apr;26(4):331-345.