Journal of Iranian Water Engineering Research

Research Article, P. 56-90

https://doi.org/10.22034/ijwer.2026.600446.1144

Multi-Step-Ahead Forecasting of Monthly Discharge of the Salmon River at Salmon in Idaho, Using SARIMA Models

Hojjat Bina[1], Mahmoud F. Maghrebi[2*]

[1]Ph.D. Student, Civil Engineering Department, Engineering Faculty, Ferdowsi University of Mashhad, Mashhad, Iran
[2]Professor, Civil Engineering Department, Engineering Faculty, Ferdowsi University of Mashhad, Mashhad, Iran

Corresponding Author's E-mail: maghrebi@um.ac.ir

Received: **22 Sep. 2026** Accepted: **18 Oct. 2026**
Revised: **11 Oct. 2026** Published: **25 Mar. 2025**

Homepage: ijwer.uoz.ac.ir

Abstract: Two SARIMA models were evaluated for fixed-origin monthly discharge forecasting at Salmon, Idaho. Training began in January 1994 and ended in December 2023 for January–December 2024 forecasts, and in December 2022 for January 2023–December 2024 forecasts, without updating with target observations. For SARIMA$(1,0,0)(1,1,2)_{12}$ and SARIMA$(1,1,2)(1,1,2)_{12}$, respectively, 12-month RMSE values were 10.709 and 9.501 m^3/s, with NSE values of 0.838 and 0.873; corresponding 24-month values were 13.480 and 17.026 m^3/s, and 0.881 and 0.810. Both models outperformed seasonal naive forecasts, but superiority over monthly climatological means was not uniform. Both underestimated the June 2023 peak in the 24-month evaluation and overestimated the June 2024 peak in the 12-month evaluation. Model structures and preprocessing settings were fixed using pre-2023 data, without using final evaluation outcomes for selection.

Keywords: Temporal validation, Seasonal differencing, Residual analysis, Benchmark models, Prediction uncertainty

Introduction: Monthly discharge forecasting requires models that represent temporal dependence and seasonality while accommodating differences between low-flow conditions and large monthly peaks. Reproducing a seasonal cycle does not establish accurate peak magnitudes or consistent performance across forecast origins. Information criteria, residual diagnostics, and predictive errors therefore address related but distinct aspects of model performance.

Alonso Brito et al. (2021) compared SARIMA and Holt–Winters models for monthly streamflow forecasting. Sheikholeslami et al. (2024) reported seasonal representation alongside limitations in forecasting exceptionally high flows in the Chalous River. Costa et al. (2023) found improved performance when precipitation was incorporated into SARIMAX. Studies of SARIMA–Prophet and STL–SARIMA also discussed limitations under high-flow conditions (Kenyi and Yamamoto, 2024; Kang et al., 2025). These findings motivate evaluation across flow conditions but do not establish transferable model rankings.

This study examines fixed-origin forecasting at the Salmon River station in Idaho. Its purpose is to compare two reference SARIMA structures, assess whether diagnostic advantages coincide with lower forecasting errors, and quantify performance against simple benchmarks. The empirical contribution concerns contrasting evaluation periods and flow conditions, rather than a new forecasting method (Figs. 5 and 6; Tables 10–14).

Methodology: The dataset comprised 372 monthly mean discharge observations from January 1994 to December 2024 at USGS station 13302500. Discharge was transformed as ln(Q+1) (Eq. (1)). ADF tests and ACF/PACF plots were used to examine the transformed series before and after seasonal-only and combined differencing (Table 3; Figs. 2 and 3). Models were estimated by maximum likelihood in MATLAB. AIC- and BIC-based rules (Eqs. (8) and (9)) were compared using expanding-window validation at December 2018–2020 origins, with 12- and 24-month horizons and all targets preceding 2023 (Table 5). Simpler structures were evaluated (Table 6).

The reference models were SARIMA(1,0,0)(1,1,2)$_{12}$ and SARIMA(1,1,2)(1,1,2)$_{12}$, with constants fixed at zero. Structures and preprocessing settings were fixed using pre-2023 data; full-record diagnostics did not inform selection or revision. Parameter estimates and residual diagnostics appear in Tables 7 and 9 and Fig. 4.

Training ended in December 2023 for January–December 2024 forecasts and December 2022 for January 2023–December 2024 forecasts (Table 1). Forecasts used fixed origins without updating with target observations. Point forecasts were back-transformed without mean bias correction and truncated at zero (Eq. (17)). Pointwise 95% intervals were transformed from the logarithmic scale, with negative lower bounds set to zero (Eqs. (18)–(21)). Performance was assessed against seasonal naive and monthly climatological forecasts using RMSE, MAE, SMAPE, and NSE (Eqs. (12)–(15); Table 10). Training-period quartiles defined flow groups (Table 2); volume bias accounted for month lengths (Eq. (16); Table 13). Both models were refitted through December 2024 for 2025–2026 forecasts (Fig. 7).

Results and Discussion: ADF results for the log-transformed series depended on the test specification. After seasonal-only and combined differencing, the unit-root null was rejected under all examined specifications, but this did not establish the necessity of both operators (Table 3; Figs. 2 and 3). The reference models also differed in nonseasonal moving-average orders; performance differences cannot therefore be attributed solely to nonseasonal differencing.

The seasonal-only reference model had the lowest aggregate validation error among the examined candidates in its group, whereas several simpler structures outperformed the combined-differencing reference model (Table 6). Adjusted Ljung–Box p-values at maximum lags 12 and 24 were 0.030 and 0.065 for the former and 0.244 and 0.250 for the latter (Table 9). Non-rejection does not prove residual independence; tail departures from normality remained visible (Fig. 4).

For the seasonal-only and combined-differencing models, respectively, 12-month RMSE values were 10.709 and 9.501 m³/s, with NSE values of 0.838 and 0.873. Over 24 months, RMSE values were 13.480 and 17.026 m³/s, and NSE values were 0.881 and 0.810. Both models outperformed seasonal naive forecasts across all four metrics. However, over 24 months, the combined-differencing model had higher RMSE and lower NSE than monthly climatological forecasts, despite lower MAE and SMAPE (Eqs. (12) and (15); Table 10).

Errors in 2023 contributed 83.63% and 91.22% of total squared errors over 24 months, respectively. For identical 2024 observations, both models had lower RMSE from December 2022 than from December 2023. Thus, error differences cannot be attributed solely to horizon; training samples and fitted parameters also differed (Tables 11 and 12).

Both models underestimated the June 2023 peak over 24 months and overestimated the June 2024 peak over 12 months (Figs. 5 and 6). This distinction between seasonal timing and peak magnitude is consistent with reported limitations (Sheikholeslami et al., 2024; Kenyi and Yamamoto, 2024; Kang et al., 2025). Over 24 months, the seasonal-only model performed better in high-flow months and the combined-differencing model in low-flow months (Table 13). Direct inverse transformation yields a conditional median under Gaussian log-scale forecasts before nonnegative truncation; its contribution to peak errors was not isolated (Eq. (17)).

All observations fell within the pointwise 95% intervals. Mean widths were 62.893 and 60.726 m³/s over 12 months and 64.749 and 59.079 m³/s over 24 months (Eqs. (18)–(21); Table 14). Complete coverage over short, overlapping periods does not establish long-term calibration. Forecasts for 2025–2026 remain unverified conditional projections (Fig. 7).

Conclusion: Neither reference model was uniformly preferable across diagnostics, evaluation periods, and flow conditions. The combined-differencing model performed better over 12 months, whereas the seasonal-only model performed better over 24 months. Both exceeded seasonal naive performance, but neither established universal superiority over simple benchmarks. Peak errors and contrasting low-flow and high-flow results demonstrate why aggregate metrics should be interpreted alongside relative errors and volume bias. Larger 24-month errors were associated with the composition of the target period and cannot be attributed solely to forecast horizon. Final evaluation outcomes were not used to select or revise model structures or preprocessing settings. Further evaluation across periods and stations is needed to assess robustness. Monthly results do not establish operational reliability or instantaneous flood-peak forecasting capability.

© 2025 University of Zabol, Zabol, Iran.

This is an open access article distributed under the terms and conditions of the Creative Commons Attribution 4.0 International (CC BY 4.0 license) (http://creativecommons.org/licenses/by/4.0/

نشریه علمی
پژوهش‌های مهندسی آب ایران

مقاله پژوهشی، ص ۹۰-۵۷
https://doi.org/10.22034/ijwer.2026.600446.1144

پیش‌بینی چندگام به جلو دبی ماهانه رودخانه سالمون در ایستگاه سالمون در آیداهو، با مدل‌های ساریما

حجت بینا[۱]، محمود فغفور مغربی[۲]*

[۱]دانشجوی دکتری، گروه عمران، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران
[۲]استاد، گروه عمران، دانشکده مهندسی، دانشگاه فردوسی مشهد، مشهد، ایران

پست الکترونیکی نگارنده مسئول: maghrebi@um.ac.ir

تاریخ دریافت: ۱۴۰۵/۰۶/۱۱ تاریخ پذیرش: ۱۴۰۵/۰۷/۲۶
تاریخ ویرایش: ۱۴۰۵/۰۷/۱۹ تاریخ چاپ: ۱۴۰۴/۰۱/۰۵
وبگاه نشریه: ijwer.uoz.ac.ir

چکیده: هدف پژوهش، مقایسه پیش‌بینی دبی ماهانه از مبدأ ثابت و بررسی محدودیت‌های دو ساختار ساریما[۱] در ایستگاه سالمون، آیداهو، بود. تعداد ۳۷۲ مشاهده از ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴، پس از تبدیل لگاریتمی، با مدل‌های $SARIMA(1,0,0)(1,1,2)_{12}$ و $SARIMA(1,1,2)(1,1,2)_{12}$ بررسی شدند. ارزیابی ساختارها بر معیارهای اطلاعاتی، تشخیص باقیمانده‌ها و اعتبارسنجی زمانی با مبدأهای دسامبر ۲۰۱۸، ۲۰۱۹ و ۲۰۲۰ استوار بود. آموزش از ژانویه ۱۹۹۴ آغاز شد و برای ارزیابی‌های ۱۲ و ۲۴ماهه به‌ترتیب تا دسامبر ۲۰۲۳ و دسامبر ۲۰۲۲ ادامه یافت؛ دوره‌های هدف به‌ترتیب ژانویه تا دسامبر ۲۰۲۴ و ژانویه ۲۰۲۳ تا دسامبر ۲۰۲۴ بودند. پیش‌بینی‌ها بدون به‌روزرسانی با مشاهدات هدف تولید شدند. ساختار و تنظیمات مدل‌ها با داده‌های پیش از ۲۰۲۳ تثبیت شدند و داده‌های هدف در انتخاب مدل‌ها دخالت نداشتند. تشخیص مطلوب‌تر باقیمانده‌ها الزاما با کمترین خطای اعتبارسنجی همراه نبود. در ارزیابی ۱۲ماهه، RMSE مدل‌های نخست و دوم به‌ترتیب ۱۰/۷۰۹ و ۹/۵۰۱ مترمکعب بر ثانیه و NSEآن‌ها ۰/۸۳۸ و ۰/۸۷۳ بود؛ در ارزیابی ۲۴ماهه، مقادیر متناظر RMSE برابر با ۱۳/۴۸۰ و ۱۷/۰۲۶ و NSE برابر با ۰/۸۸۱ و ۰/۸۱۰ بودند. هر دو مدل از روش ساده فصلی بهتر عمل کردند، اما برتری نسبت به میانگین ماهانه بلندمدت یکنواخت نبود. هر دو مدل قله ژوئن ۲۰۲۳ را در ارزیابی ۲۴ماهه کم‌برآورد و قله ژوئن ۲۰۲۴ را در ارزیابی ۱۲ماهه بیش‌برآورد کردند. بنابراین، برتری مدل‌ها به دوره و معیار ارزیابی وابسته بود و تایید پایداری عملکرد به ارزیابی مستقل در دوره‌ها و ایستگاه‌های دیگر نیاز دارد.

کلیدواژه‌ها: اعتبارسنجی زمانی، تفاضل‌گیری فصلی، تحلیل باقیمانده‌ها، مدل‌های مبنا، عدم‌قطعیت پیش‌بینی

۱- مقدمه

پیش‌بینی دبی رودخانه در برنامه‌ریزی تامین آب، بهره‌برداری از مخازن، تولید انرژی و ارزیابی شرایط کم‌آبی و پُرآبی کاربرد دارد (Kumar et al., 2023; Thota et al., 2024). در پیش‌بینی دبی ماهانه، انتخاب متغیرهای ورودی و وقفه‌های زمانی آن‌ها از تصمیم‌های مهم مدل‌سازی است. Chu و همکاران (۲۰۲۴) نقش وقفه متغیرهای ورودی را در

[1] SARIMA

پیش‌بینی جریان ماهانه بررسی کردند و Nakhaeiو همکاران (۲۰۲۴) نیز عملکرد شبکه‌های عصبی بازگشتی (RNN) و شبکه‌های حافظه بلند کوتاه‌مدت (LSTM)را با ترکیب‌ها و وقفه‌های مختلف ورودی مقایسه کردند. این یافته‌ها ضرورت توجه به اطلاعات ورودی و شرایط هر مطالعه را در تفسیر عملکرد مدل‌ها مطرح می‌کنند.

برای پیش‌بینی جریان، رویکردهای فرایندمحور، مدل‌های آماری سری زمانی و روش‌های یادگیری ماشین به کار می‌روند. مدل‌های فرایندمحور بر بازنمایی فرایندهای هیدرولوژیکی حوضه تکیه دارند؛ برای نمونه،Islam و همکاران (۲۰۲۳) عملکرد SWAT و جنگل تصادفی را در حوضه‌ای با غلبه ذوب برف مقایسه کردند. مدل‌های آماری سری زمانی از وابستگی‌های موجود در سوابق مشاهداتی استفاده می‌کنند و روش‌های یادگیری ماشین امکان بهره‌گیری از ورودی‌های متعدد و بازنمایی روابط غیرخطی را فراهم می‌سازند (Box et al., 2015; Kumar et al., 2023). در میان این رویکردها، مدل‌های خانواده آریما [1] چارچوب مشخصی برای شناسایی ساختار، برآورد پارامترها و بررسی باقیمانده‌ها دارند (Box et al., 2015) و در مطالعات پیش‌بینی جریان ماهانه نیز به‌عنوان روش پیش‌بینی یا مبنای مقایسه به کار رفته‌اند (Costa et al., 2023; Thota et al., 2024).

مدل خودرگرسیونی میانگین متحرک یکپارچه، آریما، با ترکیب مولفه‌های خودرگرسیونی، تفاضل‌گیری و میانگین متحرک، وابستگی زمانی سری را مدل‌سازی می‌کند. صورت فصلی آن، ساریما، امکان لحاظ‌کردن مولفه‌های خودرگرسیونی و میانگین متحرک فصلی و در صورت نیاز، تفاضل‌گیری فصلی را فراهم می‌سازد (Box et al., 2015) . این ساختار در مدل‌سازی دبی ماهانه دارای الگوی فصلی به کار رفته است(Costa et al., 2023) . با این حال، وجود الگوی فصلی به‌تنهایی ضرورت تفاضل‌گیری فصلی یا کفایت پیش‌بینی مدل را نشان نمی‌دهد.

کاربرد ساریما در پیش‌بینی دبی ماهانه در مطالعات مختلف بررسی شده است. Costa و همکاران (۲۰۲۳) مدل‌های ساریما و ساریماکس را در یکی از زیرحوضه‌های رودخانه پارانایبا [2] در برزیل، برای پیش‌بینی دبی متوسط ماهانه سال ۲۰۱۶ در افق‌های ۳، ۶، ۹ و ۱۲ماهه مقایسه کردند. افزودن بارش به‌عنوان متغیر برون‌زا، شاخص‌های ارزیابی ساریماکس را نسبت به ساریما بهبود داد؛ بااین‌حال، کم‌برآوردی قله‌ها در دوره آزمون و خطا در ماه‌های خشک همچنان گزارش شد. Sheikholeslami و همکاران (۲۰۲۴) نیز مدل ساریما را با استفاده از داده‌های ماهانه سال‌های ۲۰۰۶ تا ۲۰۲۳ ایستگاه پل زغال در رودخانه چالوس بررسی کردند. نویسندگان گزارش کردند که مدل، با وجود بازنمایی تغییرات فصلی جریان، در پیش‌بینی دبی‌های بسیار بزرگ با محدودیت مواجه بود. Alonso Britoو همکاران (۲۰۲۱) نیز با استفاده از دبی متوسط ماهانه حوضه وی آنیورساریو [3] در کوبا طی سال‌های ۱۹۷۱ تا ۱۹۹۰، مدل‌های ساریما و هولت-وینترز [4] را در اعتبارسنجی با مبدأهای پیش‌رونده و مجموعه آموزش گسترش‌یابنده مقایسه کردند. پیش‌بینی‌ها در افق‌های ۱۲ و ۲۴ماهه ارزیابی شدند. هولت-وینترز با مجموعه‌های آموزش کوتاه‌تر عملکرد بهتری داشت؛ با افزایش طول آموزش، عملکرد دو روش در افق ۱۲ماهه نزدیک شد و ساریما در افق ۲۴ماهه و بازنمایی جریان‌های متغیر فصل مرطوب برتری نسبی نشان داد. با وجود این، هر دو روش در پیش‌بینی برخی قله‌ها و جریان‌های بسیار کم خطا داشتند.

ساریما در مدل‌های ترکیبی پیش‌بینی جریان نیز به کار رفته است. Danandeh Mehr و همکاران (۲۰۲۲) عملکرد برنامه‌ریزی ژنتیک و ساریما را برای پیش‌بینی یک‌گام‌به‌جلوی جریان در مقیاس‌های روزانه، هفتگی و ماهانه در سرشاخه‌های سامانه رودخانه‌ای-دریاچه‌ای اولو یوکی [5] در فنلاند بررسی کردند. مدل ترکیبی جی‌پی-ساریما [6] خطای پیش‌بینی ماهانه را نسبت به دو مدل مستقل کاهش داد، اما کم‌برآوردی برخی جریان‌های بزرگ باقی ماند. Kenyi و Yamamoto (۲۰۲۴) در رودخانه

[1] ARIMA
[2] Paranaíba
[3] V Aniversario
[4] Holt–Winters
[5] Oulujoki
[6] GP–SARIMA

سوبات[1] در سودان جنوبی، عملکرد مدل ترکیبی ساریما-پروفت[2] را با دو مدل مستقل مقایسه کردند. مدل ترکیبی MAE و RMSE کمتر و ضریب تعیین[3](R^2) بیشتری داشت؛ درحالی‌که هر دو مدل مستقل در بازنمایی مقادیر حدی با محدودیت مواجه بودند. Kang و همکاران (۲۰۲۵) نیز برای پیش‌بینی جریان ورودی ایستگاه برق‌آبی اِرتان[4] در حوضه رودخانه یالونگ[5]، مدل اس‌تی‌ال-ساریما[6] را توسعه دادند. در این روش، سری جریان به مولفه‌های روند، فصلی و باقیمانده تفکیک و پیش‌بینی نهایی از ترکیب پیش‌بینی مولفه‌ها حاصل شد. خطای مدل ترکیبی از مدل‌های مستقل ساریما و آریما کمتر بود، ولی اختلاف میان مقادیر بیشینه پیش‌بینی‌شده و مشاهده‌شده برطرف نشد.

Kamal و همکاران (۲۰۲۴) مدل آریما را برای پیش‌بینی دبی ماهانه رودخانه آتاباسکا[7] در سه ایستگاه هینتون[8]، آتاباسکا و فورت مک‌موری[9] به کار بردند و تفاوت خطای پیش‌بینی میان ایستگاه‌ها را گزارش کردند. Thota و همکاران (۲۰۲۴) نیز مدل‌های جنگل تصادفی، ال‌اس‌تی‌ام[10]،جی‌آریو[11] و ساریما را برای پیش‌بینی جریان ماهانه در لیز فری[12] در حوضه بالادست کلرادو[13]، با داده‌های سال‌های ۱۹۹۱ تا ۲۰۲۰ مقایسه کردند. آموزش با داده‌های ۱۹۹۱ تا ۲۰۱۴ و آزمون با داده‌های ۲۰۱۵ تا ۲۰۲۰ انجام شد. عملکرد مدل‌ها با استفاده از سابقه جریان و با افزودن بارش، دما و معادل آبی برف بررسی شد. بهترین عملکرد گزارش‌شده به جنگل تصادفی با ورودی ۲۴ماهه، خروجی ۱۲ماهه و استفاده از همه متغیرهای هواشناختی تعلق داشت. در تفسیر این نتایج، باید طول شش‌ساله مجموعه آزمون را از افق پیش‌بینی ۱۲ماهه متمایز دانست.

در پیش‌بینی کوتاه‌مدت روزانه نیز ترکیب مدل‌ها و تنظیم پارامترهای آن‌ها بررسی شده است. Martinho و همکاران (۲۰۲۳) چند الگوریتم یادگیری ماشین را با روش‌های بهینه‌سازی فرا ابتکاری برای پیش‌بینی جریان در افق‌های ۱، ۳، ۵ و ۷روزه تلفیق کردند. در مقایسه آن‌ها، الگوریتم ازدحام ذرات برای تنظیم پارامترها عملکرد بهتری از الگوریتم ژنتیک و تکامل تفاضلی داشت. Di Nunno و همکاران (۲۰۲۳) نیز مدلی ترکیبی شامل نارکس[14]، پرسپترون چندلایه و جنگل تصادفی را برای پیش‌بینی جریان تا هفت روز آینده در ۱۸ رودخانه بریتانیا ارزیابی کردند و در مجموع، عملکرد بهتری نسبت به مدل‌های منفرد گزارش دادند. این نتایج به مقیاس روزانه و افق‌های کوتاه مربوط‌اند و نباید مستقیما به پیش‌بینی چندماهه تعمیم داده شوند.

Akkala و همکاران (۲۰۲۵) با استفاده از داده‌های ۲۰ ایستگاه حوضه بالادست کلرادو، عملکرد شبکه عصبی گرافی مکانی-زمانی را در کنار مدل‌های آماری و یادگیری ماشین بررسی کردند. در حالت چندمتغیره، این شبکه در ایستگاه لیز فری بالاترین دقت را میان مدل‌های بررسی‌شده داشت و افزودن معادل آبی برف با کاهش خطای پیش‌بینی همراه بود.

مطالعات مرورشده از نظر اطلاعات ورودی، افق پیش‌بینی و طرح ارزیابی متفاوتاند؛ ازاین‌رو، مقایسه عملکرد مدل‌ها مستلزم توجه به این تفاوت‌هاست. مقایسه با روش ساده فصلی نیز مشخص می‌کند که مدل تا چه اندازه از پیش‌بینی مبتنی بر آخرین دبی مشاهده‌شده همان ماه تقویمی در دوره آموزش فراتر می‌رود.

مسئله پژوهش حاضر، بررسی همسویی شواهد حاصل از معیارهای اطلاعاتی و تشخیص باقیمانده‌ها با عملکرد پیش‌بینی ساریما و ارزیابی محدودیت‌های آن در شرایط مختلف جریان است. برای این منظور، ۳۷۲ مشاهده ماهانه ایستگاه رودخانه سالمون در شهر سالمون، آیداهو، با شناسه USGS 13302500، طی ژانویه ۱۹۹۴ تا دسامبر

[1] Sobat
[2] SARIMA–Prophet
[3] Coefficient of determination
[4] Ertan
[5] Yalong
[6] STL–SARIMA
[7] Athabasca
[8] Hinton
[9] Fort McMurray
[10] LSTM
[11] GRU
[12] Lees Ferry
[13] Colorado
[14] NARX

۲۰۲۴ بررسی شد. در ارزیابی‌های پایانی ۱۲ و ۲۴ماهه، پیش‌بینی‌ها به‌صورت تک‌متغیره و چندگام‌به‌جلو از مبدأ ثابت تولید شدند؛ به این معنا که مشاهدات جدید دوره هدف وارد فرایند پیش‌بینی نشدند. این طرح، عملکرد مدل را با اتکا به سابقه دبی ارزیابی می‌کند. بارش، دما و معادل آبی برف مستقیما وارد مدل نمی‌شوند و با توجه به اجرا نشدن مدل چندمتغیره، سهم نبود این ورودی‌ها در خطاهای مشاهده‌شده قابل تعیین نیست.

دو گروه ساختارساریما، شامل تفاضل‌گیری فصلی به‌تنهایی و تفاضل‌گیری فصلی همراه با تفاضل‌گیری غیرفصلی، بررسی شدند. قواعد انتخاب مبتنی بر معیارهای اطلاعاتی در سه مبدأ دسامبر ۲۰۱۸، ۲۰۱۹ و ۲۰۲۰ و افق‌های ۱۲ و ۲۴ماهه اعتبارسنجی شدند. تمام مشاهدات هدف این اعتبارسنجی به پیش از سال ۲۰۲۳ تعلق داشتند. ساختارهای ساده‌تر نیز در همین مبدأها و افق‌ها مقایسه شدند. عملکرد دو ساختار مورد مطالعه در دوره‌های پایانی ۱۲ و ۲۴ماهه نیز با روش ساده فصلی و میانگین ماهانه بلندمدت مقایسه شد. ساختار و تنظیمات مدل‌ها پیش از ارزیابی دوره‌های پایانی تثبیت شدند و مشاهدات یا خطاهای این دوره‌ها در انتخاب یا بازنگری مدل‌ها دخالت نداشتند. برای هر دو ساختار مورد مطالعه، خطاها به تفکیک سال، ماه و شرایط کم‌آبی و پُرآبی، همراه با سوگیری حجمی، پوشش تجربی و پهنای بازه‌های پیش‌بینی بررسی شدند.

سهم این پژوهش، مستندسازی رابطه میان معیار انتخاب مدل و عملکرد پیش‌بینی، و کمی‌سازی محدودیت‌های ساریما در ایستگاه و دوره‌های بررسی‌شده است. تحلیل‌ها بر سه پرسش متمرکزند: شواهد معیارهای اطلاعاتی و تشخیص باقیمانده‌ها تا چه اندازه با خطای اعتبارسنجی همسو هستند؛ دو ساختار مورد مطالعه نسبت به روش‌های مبنا چه عملکردی دارند؛ و تحلیل تفکیکی خطا و بازه‌های پیش‌بینی چه اطلاعاتی فراتر از شاخص‌های تجمیعی فراهم می‌کند. دامنه استنباط به همین ایستگاه و طرح ارزیابی محدود است و نتایج، مبنایی برای ادعای برتری عمومی ساریما یا تایید کاربرد عملیاتی آن فراهم نمی‌کنند.

۲- روششناسی

۲-۱- منطقه مورد مطالعه و داده‌ها

داده‌های پژوهش از ایستگاه رودخانه سالمون در شهر سالمون، ایالت آیداهو، متعلق به سازمان زمین‌شناسی ایالات متحده آمریکا (USGS)، با شناسه ۱۳۳۰۲۵۰۰، دریافت شدند. این ایستگاه در عرض جغرافیایی ۴۵/۱۸۳۶۱۱ درجه شمالی و طول جغرافیایی ۱۱۳/۸۹۵۲۷۸ درجه غربی، بر مبنای دستگاه مرجع WGS84، قرار دارد. مساحت حوضه بالادست ایستگاه ۳۷۳۷ مایل مربع، معادل تقریبی ۹۶۷۸/۸ کیلومترمربع، است. داده‌های دریافت‌شده شامل دبی با کد پارامتر ۰۰۰۶۰ و آماره میانگین روزانه با کد ۰۰۰۰۳، در واحد فوت‌مکعب بر ثانیه بودند.

دوره مطالعه از ۱ ژانویه ۱۹۹۴ تا ۳۱ دسامبر ۲۰۲۴ را دربر گرفت و شامل ۱۱۳۲۳ رکورد یکتای دبی متوسط روزانه بود. بررسی پیوستگی زمانی و مقادیر داده‌ها نشان داد که روز فاقد رکورد، مقدار خالی یا ماه ناقص وجود نداشت؛ بنابراین، حذف ماه یا جایگزینی داده مفقود انجام نشد. تمام رکوردهای دوره مطالعه دارای وضعیت تاییدشده سازمان زمین‌شناسی ایالات متحده آمریکا بودند. از میان آن‌ها، ۱۵۸۷ رکورد نشان برآوردی[1] داشتند که بدون تغییر در محاسبات حفظ شدند.

دبی متوسط هر ماه از میانگین حسابی دبی‌های متوسط روزانه همان ماه، با استفاده از تمام روزهای ماه و با لحاظ سال‌های کبیسه، محاسبه شد. مقادیر حاصل با ضریب ۰/۰۲۸۳۱۶۸۵ از فوت‌مکعب بر ثانیه به مترمکعب بر ثانیه تبدیل شدند. بدین‌ترتیب، سری نهایی شامل ۳۷۲ مقدار دبی متوسط ماهانه از ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴ به دست آمد.

پیش از مدل‌سازی، برای فشرده‌کردن دامنه تغییرات دبی و کاهش اثر مقادیر بزرگ بر برازش در مقیاس تبدیل‌شده،

[1] ESTIMATED

تبدیل لگاریتم طبیعی زیر اعمال شد:

$$y_t=ln(Q_t+1) \quad (1)$$

که در آن، Q_t مقدار عددی دبی متوسط ماهانه در زمان (t)، بر حسب مترمکعب‌برثانیه، و y_t مقدار تبدیل‌شده آن است. افزودن عدد ۱، محاسبه تبدیل را برای دبی صفر نیز امکان‌پذیر می‌کند؛ بااین‌حال، داده‌های ماهانه این پژوهش فاقد مقدار صفر یا منفی بودند و حداقل دبی برابر با ۱۱/۳۸۵ مترمکعب بر ثانیه بود. بنابراین، هیچ مشاهده‌ای به دلیل صفر بودن حذف یا جایگزین نشد. تبدیل لگاریتمی به‌تنهایی تضمین‌کننده تثبیت واریانس یا نرمال‌شدن توزیع داده‌ها نیست. مدل‌سازی بر مبنای سری تبدیل‌شده انجام شد و پیش‌بینی‌ها مطابق روش تشریح‌شده در بخش ۲-۱۱ به مقیاس اصلی دبی بازگردانده شدند.

۲-۲- بررسی مانایی و تعیین ساختار تفاضل‌گیری

در روش باکس-جنکینز[1]، ساختار خودرگرسیونی و میانگین متحرک بر سری مانا یا سری حاصل از تفاضل‌گیری مدل‌سازی می‌شود؛ ازاین‌رو، بررسی مانایی و تعیین مرتبه تفاضل‌گیری از مراحل شناسایی مدل است (Box et al., 2015).

برای بررسی وجود ریشه واحد، آزمون دیکی-فولر تعمیم‌یافته (ADF)[2] با تابع adftest در نرم‌افزار متلب و در سطح معنی‌داری ۵ درصد اجرا شد. فرض صفر این آزمون، وجود ریشه واحد مورد آزمون است و مقدار احتمال کمتر از ۰/۰۵ ملاک رد آن قرار گرفت (Dickey and Fuller, 1981). اجرای اولیه با مشخصات AR، بدون ثابت و روند خطی و با صفر وقفه افزوده انجام شد. برای بررسی حساسیت نتایج، آزمون با مشخصات ARD، شامل ثابت و بدون روند خطی، در سه حالت صفر، ۱۲ و ۲۴ وقفه افزوده تکرار شد. تعداد وقفه‌ها برای این تحلیل ثابت تعیین شد و حاصل انتخاب خودکار با معیارهای اطلاعاتی نبود.

آزمون بر سه سری شامل دبی لگاریتمی بدون تفاضل‌گیری، سری دارای یک مرتبه تفاضل‌گیری فصلی و سری دارای یک مرتبه تفاضل‌گیری غیرفصلی و فصلی اجرا شد. دوره داده، تعداد مشاهدات سری، تعداد مشاهدات موثر رگرسیون، مشخصات آزمون، آماره، مقدار بحرانی و مقدار احتمال گزارش شدند. آزمون‌های گزارش‌شده در این بخش بر کل دوره ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴ اجرا شدند و نقش توصیفی و تحلیل حساسیت داشتند. نتایج این آزمون‌ها در انتخاب یا تغییر ساختار و تنظیمات مدل‌های ارزیابی‌شده دخالت نداشتند؛ این تصمیم‌ها با استفاده از داده‌های پیش از ۲۰۲۳ و اعتبارسنجی زمانی تثبیت شده بودند.

رد فرض صفر، شواهدی علیه ریشه واحد مورد آزمون فراهم می‌کند، اما به‌تنهایی ضرورت تفاضل‌گیری فصلی یا مرتبه آن را تعیین نمی‌کند. همچنین، ردنشدن فرض صفر به معنی اثبات نامانایی نیست. ازاین‌رو، نتایج آزمون در کنار نمودارهای خودهمبستگی و خودهمبستگی جزئی و ارزیابی پیش‌بینی ساختارهای رقیب بررسی شدند.

تفاضل غیرفصلی مرتبه اول به صورت زیر تعریف شد:

$$\nabla y_t=(1-B)y_t=y_t-y_{t-1} \quad (2)$$

و تفاضل فصلی مرتبه اول با دوره ۱۲ ماهه به شکل زیر در نظر گرفته شد:

$$\nabla_{12} y_t=(1-B^{12})y_t=y_t-y_{t-12} \quad (3)$$

که در آن، y_t سری لگاریتمی دبی، B عملگر وقفه ($By_t=y_{t-1}$)، ∇ عملگر تفاضل‌گیری غیرفصلی مرتبه اول و ∇_{12} عملگر تفاضل‌گیری فصلی مرتبه اول با دوره ۱۲ ماهه است.

در ساختار دارای یک مرتبه تفاضل‌گیری غیرفصلی و یک مرتبه تفاضل‌گیری فصلی، سری حاصل از اعمال هم‌زمان دو تفاضل به‌صورت زیر تعریف می‌شود:

$$\omega_t=(1-B)(1-B^{12})y_t \quad (4)$$

[1] Box–Jenkins method

[2] Augmented Dickey–Fuller test (ADF)

یا به‌صورت بسط‌یافته:

$$\omega_t=y_t-y_{t-1}-y_{t-12}+y_{t-13} \quad (5)$$

که در آن، ω_t مقدار سری لگاریتمی پس از اعمال هم‌زمان یک مرتبه تفاضل‌گیری غیرفصلی و یک مرتبه تفاضل‌گیری فصلی با دوره ۱۲ ماهه است. دو گروه با مرتبه‌های (d=0,D=1) و (d=1,D=1) بررسی شدند. برای اجرای آزمون و ترسیم نمودارهای تشخیصی، تفاضل‌ها مستقیما از سری لگاریتمی محاسبه شدند. در برازش مدل، سری لگاریتمی بدون تفاضل‌گیری دستی به تابع برآورد داده شد و تفاضل‌گیری از طریق مشخصات مدل اعمال شد.

در مرحله اعتبارسنجی زمانی، معیارهای اطلاعاتی آکائیک و بیزی برای غربال مدل‌های هر گروه با مرتبه تفاضل‌گیری یکسان و بر داده‌های آموزش مشترک به کار رفتند. مقایسه عملکرد دو گروه بر خطای اعتبارسنجی زمانی استوار بود و نتایج تشخیص باقیمانده‌ها نیز جداگانه بررسی شدند. وجود الگوی فصلی، مبنای بررسی ساختارهای فصلی بود و به‌عنوان اثبات وجود ریشه واحد فصلی یا ضرورت قطعی (D=1) تلقی نشد.

۲-۳- شناسایی ساختار وابستگی زمانی

ساختار وابستگی زمانی سری با استفاده از تابع خودهمبستگی[1] (ACF) و تابع خودهمبستگی جزئی[2] (PACF) بررسی شد. تابع خودهمبستگی، همبستگی خطی میان مقادیر سری در وقفه‌های مختلف را نشان می‌دهد؛ تابع خودهمبستگی جزئی نیز این همبستگی را پس از کنترل اثر خطی وقفه‌های میانی بیان می‌کند. این توابع در روش باکس-جنکینز برای شناسایی اولیه مرتبه مولفه‌های خودرگرسیونی و میانگین متحرک به کار می‌روند (Box et al., 2015).

نمودارهای خودهمبستگی و خودهمبستگی جزئی برای سری لگاریتمی بدون تفاضل‌گیری و دو سری حاصل از تفاضل‌گیری با مرتبه‌های (d=0,D=1) و (d=1,D=1)، تا وقفه ۴۰ماهه بررسی شدند. دامنه وقفه‌ها در همه نمودارها یکسان بود تا وابستگی‌های کوتاه‌مدت و فصلی پیش و پس از تفاضل‌گیری قابل مقایسه باشند. در تفسیر نمودارها، نحوه کاهش خودهمبستگی، باقی‌ماندن همبستگی در وقفه‌های فصلی و ظهور همبستگی‌های منفی پس از تفاضل‌گیری بررسی شد. همبستگی‌های منفی به‌تنهایی نشانه قطعی تفاضل‌گیری بیش‌ازحد تلقی نشدند و کوتاه‌ترشدن دنباله خودهمبستگی نیز ملاک کافی برای ترجیح یک ساختار نبود. شواهد نموداری در کنار نتایج آزمون ریشه واحد و ارزیابی پیش‌بینی ساختارهای رقیب تفسیر شدند.

۲-۴- ساختار مدل‌های آریما و ساریما

برای مقایسه ساختارهای غیرفصلی و فصلی در مدل‌سازی دبی ماهانه، مدل‌های آریما و ساریما در چارچوب روش باکس-جنکینز بررسی شدند (Box et al., 2015). در مدل آریما با ساختار ARIMA(p, d, q)، پارامترهای p، d و q به‌ترتیب مرتبه خودرگرسیونی، مرتبه تفاضل‌گیری غیرفصلی و مرتبه میانگین متحرک هستند. رابطه عمومی مدل به‌صورت زیر است:

$$\phi_p(B)\times(1-B)^d\times y_t=c+\theta_q(B)\times\varepsilon_t \quad (6)$$

که در آن، $\phi_p(B)$ و $\theta_q(B)$ به‌ترتیب چندجمله‌ای‌های خودرگرسیونی و میانگین متحرک غیرفصلی، c ثابت مدل و ε_t جمله نوآوری یا خطای تصادفی مدل در زمان t است که دارای میانگین صفر، واریانس ثابت و فاقد خودهمبستگی فرض می‌شود. برای لحاظ‌کردن فصل‌گرایی، ساختار عمومی ساریما به صورت SARIMA (p, d, q) (P, D, Q)$_s$ در نظر گرفته شد، که در آن P، D و Q به‌ترتیب مرتبه‌های خودرگرسیونی فصلی، تفاضل فصلی و میانگین متحرک فصلی هستند و s طول دوره فصلی است که در این مطالعه برابر با ۱۲ ماه در نظر گرفته شد. فرم عمومی مدل ساریما را می‌توان به صورت زیر نوشت:

$$\phi_p(B)\times\Phi_P(B^s)\times(1-B)^d\times(1-B^s)^D\times \quad (7)$$
$$y_t=c+\theta_q(B)\times\Theta_Q(B^s)\times\varepsilon_t$$

[1] Autocorrelation function (ACF)

[2] Partial autocorrelation function (PACF)

که در آن، $\phi_p(B)$ و $\theta_q(B)$ به‌ترتیب چندجمله‌ای‌های خودرگرسیونی و میانگین متحرک غیرفصلی و $\Phi_P(B^s)$ و $\Theta_Q(B^s)$ به‌ترتیب چندجمله‌ای‌های خودرگرسیونی و میانگین متحرک فصلی هستند. در برازش مدل‌های این پژوهش، ثابت معادله برابر صفر قرار داده شد و برآورد نشد. دامنه مرتبه‌های بررسی‌شده و نحوه غربال و مقایسه مدل‌ها در بخش انتخاب مدل تشریح می‌شود.

۲-۵- برآورد، غربال و انتخاب مدل‌های کاندید

پارامترهای مدل‌ها با تابع estimate در محیط متلب و به روش بیشینه درست‌نمایی، با فرض نوآوری‌های گاوسی، برآورد شدند. برای انتخاب ساختار، جست‌وجوی مدل‌های ساریما در دو گروه تفاضل‌گیری $d=0$ ، $D=1$ و $d=1$، $D=1$ انجام شد. در هر گروه، مرتبه‌های p، q، P و Q به‌طور مستقل از مجموعه $\{0,1,2\}$ انتخاب شدند؛ بنابراین، ۸۱ ساختار در هر گروه بررسی شد. دوره فصلی برابر ۱۲ ماه و ثابت همه مدل‌ها برابر صفر در نظر گرفته شد. ثابت به‌عنوان پارامتر آزاد برآورد نشد. برازش‌ها با تنظیمات پیش‌فرض تابع estimate انجام شدند. پذیرش عددی برازش در جست‌وجوی ساختارها به مثبت‌بودن پرچم خروج بهینه‌سازی، متناهی‌بودن لگاریتم درست‌نمایی و قابل‌محاسبه‌بودن معیارهای اطلاعاتی مشروط بود؛ خطاها و هشدارهای اجرا نیز ثبت شدند.

برای غربال ساختارها از معیارهای اطلاعاتی آکائیک[1] و بیزی[2] استفاده شد (Akaike, 1974; Schwarz, 1978).

$$AIC=-2ln(\hat{L})+2K \quad (8)$$

$$BIC=-2ln(\hat{L})+K\,ln(n) \quad (9)$$

که در آنها $\hat{L}$ مقدار بیشینه درست‌نمایی، K تعداد پارامترهای آزاد، شامل ضرایب خودرگرسیونی، میانگین متحرک و واریانس نوآوری است. با توجه به ثابت‌بودن جمله ثابت، برای ساختارهای بررسی‌شده $K=p+q+P+Q+1$ بود. جمله‌های حاصل از ضرب چندجمله‌ای‌های فصلی و غیرفصلی، پارامتر مستقل محسوب نشدند. همچنین، n برابر تعداد مشاهدات ورودی دوره آموزش بود. در این غربال‌گری، معیارهای اطلاعاتی فقط میان مدل‌های دارای مرتبه تفاضل‌گیری یکسان و برازش‌شده بر داده‌های آموزش مشترک مقایسه شدند.

برای تعیین نحوه تصمیم‌گیری هنگام اختلاف معیارهای اطلاعاتی آکائیک و بیزی، عملکرد دو قاعده انتخاب در اعتبارسنجی زمانی مقایسه شد .دوره‌های آموزش از ژانویه ۱۹۹۴ آغاز شدند و به‌ترتیب در دسامبر ۲۰۱۸، ۲۰۱۹ و ۲۰۲۰ پایان یافتند .در هر مبدأ و در هر گروه تفاضل‌گیری، ساختار دارای کمترین مقدار هر معیار صرفا بر اساس داده‌های آموزش همان مبدأ انتخاب شد و پیش‌بینی‌های آن در افق‌های ۱۲ و ۲۴ماهه ارزیابی شدند .ازاین‌رو، ساختار منتخب هر قاعده می‌توانست میان مبدأهای اعتبارسنجی متفاوت باشد. پیش‌بینی‌ها با روش تبدیل معکوس تشریح‌شده در بخش مربوط به مقیاس اصلی دبی بازگردانده شدند. امتیاز هر قاعده، میانگین حسابی شش مقدار ریشه میانگین مربعات خطا حاصل از سه مبدأ و دو افق بود؛ بنابراین، هر ترکیب مبدأ و افق وزن یکسانی داشت .قاعده دارای امتیاز کمتر ترجیح داده شد و در صورت برابری دقیق امتیازها، معیار اطلاعاتی بیزی ملاک قرار گرفت .نتایج آزمون‌های باقیمانده‌ها برای بررسی کفایت ساختارها گزارش شدند و شرط حذف خودکار مدل‌ها در این اعتبارسنجی نبودند.

برای بررسی نقش پیچیدگی، مجموعه‌ای از ساختارهای ساده‌تر با کاهش مرتبه‌های خودرگرسیونی و میانگین متحرک فصلی و غیرفصلی هر یک از دو مدل مرجع و با ثابت نگه‌داشتن مرتبه‌های تفاضل‌گیری تعریف شد. مرتبه‌های هر ساختار در طول اعتبارسنجی ثابت ماندند و پارامترهای آن در هر مبدأ، با داده‌های آموزش همان مبدأ برآورد شدند. این بررسی در همان سه مبدأ و دو افق انجام شد و با احتساب مدل مرجع، شامل ۱۲ ساختار در گروه $d=0$ و $D=1$ و ۳۶ ساختار در گروه $d=1$ و $D=1$ بود. ساختارها از نظر تعداد پارامترهای آزاد، معیارهای اطلاعاتی،

[1] Akaike information criterion (AIC)

[2] Bayesian information criterion (BIC)

خطای اعتبارسنجی و نتایج آزمون لیونگ-باکس مقایسه شدند.

۲-۶- تشخیص باقیمانده‌ها و ارزیابی کفایت مدل

کفایت تشخیصی مدل‌های بررسی‌شده با تحلیل باقیمانده‌ها در مقیاس لگاریتمی ارزیابی شد. باقیمانده‌ها با تابع infer در محیط متلب استخراج شدند و به‌صورت اختلاف مقدار مشاهده‌شده سری تبدیل‌شده و پیش‌بینی یک‌گام‌به‌جلو آن تعریف شدند:

$$e_t=y_t-\hat{y}_t \qquad (10)$$

که در آن، e_t باقیمانده مدل، y_t مقدار مشاهده‌شده و $\hat{y}_t$ مقدار برازش‌شده یا پیش‌بینی یک‌گام‌به‌جلو در زمان t است که با استفاده از اطلاعات موجود تا زمان $t-1$ محاسبه می‌شود.

برای بررسی الگوهای باقی‌مانده، نمودار زمانی، توابع خودهمبستگی و خودهمبستگی جزئی، هیستوگرام و نمودار چندک-چندک باقیمانده‌های مدل‌های مرجع بررسی شدند. نمودارهای تشخیصی اصلی از برازش مدل‌ها به کل سری تهیه شدند؛ در اعتبارسنجی زمانی، آزمون باقیمانده‌ها برای ساختار منتخب هر قاعده، صرفا با داده‌های آموزش همان مبدأ انجام شد.

برای آزمون مشترک نبود خودهمبستگی تا وقفه‌های ۱۲ و ۲۴ماهه، از آزمون لیونگ-باکس[1] استفاده شد (Ljung and Box, 1978). باقیمانده‌ها پیش از اجرای آزمون بر انحراف معیار برآوردشده نوآوری‌ها تقسیم شدند. آماره آزمون به‌صورت زیر است:

$$Q_{LB}=n(n+2)\sum_{k=1}^{m}\frac{\hat{\rho}_k^2}{n-k} \qquad (11)$$

که در آن، n تعداد باقیمانده‌های مورد استفاده در آزمون، $\hat{\rho}_k$ ضریب خودهمبستگی باقیمانده‌ها در وقفه k و m حداکثر وقفه مورد بررسی است.

برای لحاظ‌کردن برآورد پارامترهای پویای مدل، درجه آزادی توزیع تقریبی کای‌دو در مدل‌های SARIMA برابر با $(m-(p+q+P+Q))$ در نظر گرفته شد(Box et al., 2015). مرتبه‌های تفاضل‌گیری، واریانس نوآوری‌ها و ثابتِ تثبیت‌شده در صفر، در شمار پارامترهای پویای این تعدیل منظور نشدند.

فرض صفر آزمون، صفر بودن همبستگی‌های باقیمانده‌ها در وقفه‌های ۱ تا (m) است. در سطح معنی‌داری ۵ درصد، مقدار احتمال بزرگ‌تر از ۰/۰۵ به‌عنوان نبود شواهد کافی برای رد این فرض تفسیر شد؛ این نتیجه به‌تنهایی اثبات‌کننده استقلال باقیمانده‌ها یا قابلیت اتکای پیش‌بینی‌ها نیست. هیستوگرام و نمودار چندک-چندک[2] نیز برای بررسی شکل توزیع و انحراف احتمالی از توزیع نرمال به کار رفتند و مبنای تایید قطعی نرمال‌بودن خطاها قرار نگرفتند.

۲-۷- مقایسه ساختارهای غیرفصلی و فصلی

برای مقایسه تشخیصی ساختارهای غیرفصلی، مدل‌های ARIMA با مرتبه‌های p و q از مجموعه $\{0,1,2\}$ در دو گروه $d=0$ و $d=1$ بررسی شدند؛ بنابراین، در هر گروه ۹ ساختار و در مجموع ۱۸ ساختار برازش شد. مدل‌ها با ثابت صفر و فرض نوآوری‌های گاوسی، به روش بیشینه درست‌نمایی به سری $\ln(Q+1)$ در کل دوره ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴ برازش شدند. پذیرش عددی برازش به مثبت‌بودن پرچم خروج بهینه‌سازی، متناهی‌بودن لگاریتم درست‌نمایی و معیارهای اطلاعاتی و مثبت‌بودن واریانس نوآوری مشروط بود. تعداد پارامترهای آزاد برابر با $K=p+q+1$ در نظر گرفته شد و مقایسه AIC و BIC صرفا درون هر گروه با مرتبه تفاضل‌گیری یکسان انجام گرفت. باقیمانده‌های ساختارهای پیشنهادی هر معیار با آزمون لیونگ-باکس تا وقفه‌های ۱۲ و ۲۴ و درجه آزادی $m-p-q$ بررسی شدند. این تحلیل، مقایسه تشخیصی کل دوره بود و در انتخاب یا بازنگری ساختارهای پیش‌بینی ارزیابی‌شده دخالت نداشت.

برای دو ساختار SARIMA مورد مطالعه، معیارهای

[1] Ljung–Box test

[2] Quantile–quantile plot

اطلاعاتی، خطای اعتبارسنجی زمانی و نتایج تشخیص باقیمانده‌ها جداگانه بررسی شدند تا مشخص شود ترجیح مدل بر اساس هر یک از این ملاک‌ها تا چه اندازه با سایر ملاک‌ها سازگار است. هر دو ساختار در تحلیل حفظ شدند و هیچ‌یک صرفا بر مبنای معیارهای اطلاعاتی یا آزمون باقیمانده‌ها برتر تلقی نشد. از آنجا که این ساختارها علاوه بر مرتبه تفاضل‌گیری معمولی، در مرتبه میانگین متحرک غیرفصلی نیز تفاوت دارند، اختلاف عملکرد آن‌ها صرفا به اثر تفاضل‌گیری معمولی نسبت داده نشد.

۲-۸- ارزیابی پیش‌بینی در دوره‌های پایانی

عملکرد دو ساختار مورد مطالعه در دوره‌های ژانویه تا دسامبر ۲۰۲۴ و ژانویه ۲۰۲۳ تا دسامبر ۲۰۲۴ بررسی شد. این دوره‌ها از اعتبارسنجی زمانی با مبدأهای دسامبر ۲۰۱۸، ۲۰۱۹ و ۲۰۲۰ متمایز بودند. در اعتبارسنجی زمانی، پنجره آموزش از ژانویه ۱۹۹۴ تا هر مبدأ گسترش یافت و تمام مشاهدات هدف به پیش از ژانویه ۲۰۲۳ تعلق داشتند. جزئیات مقایسه قواعد انتخاب در بخش ۵-۲ ارائه شده است. تبدیل لگاریتمی، روش تبدیل معکوس و سایر تنظیمات مشترک در اجراهای اعتبارسنجی یکسان نگه داشته شدند.

دوره‌های آموزش و هدف در جدول ۱ ارائه شده‌اند. برای پیش‌بینی ۱۲ماهه سال ۲۰۲۴، پارامترهای هر مدل با داده‌های ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۳ برآورد شدند. برای پیش‌بینی ۲۴ماهه سال‌های ۲۰۲۴-۲۰۲۳، دوره آموزش در دسامبر ۲۰۲۲ پایان یافت. در هر سناریو، مدل یک‌بار برازش شد و پیش‌بینی‌های چندگام‌به‌جلو از مبدأ ثابت، بدون استفاده از مشاهدات دوره هدف یا برآورد مجدد پارامترها در طول آن دوره، تولید شدند.

انتخاب ساختارها و تثبیت تنظیمات پیش‌پردازش صرفا با استفاده از داده‌های پیش از ۲۰۲۳ و پیش از بررسی عملکرد در دوره‌های هدف انجام شد. مشاهدات و خطاهای دوره‌های پایانی در انتخاب ساختار، تغییر پیش‌پردازش یا بازنگری مدل‌ها استفاده نشدند. در هر سناریو، پارامترهای ساختار تثبیت‌شده فقط با داده‌های موجود تا مبدأ پیش‌بینی برآورد شدند. برای ارزیابی ۲۴ماهه، آموزش تا دسامبر ۲۰۲۲ و برای ارزیابی ۱۲ماهه، آموزش تا دسامبر ۲۰۲۳ ادامه یافت. استفاده از داده‌های سال ۲۰۲۳ در سناریوی دوم به برآورد مجدد پارامترها محدود بود و ساختار مدل تغییر نکرد. پیش‌بینی‌ها از مبدأ ثابت و بدون ورود مشاهدات هدف تولید شدند. دو ارزیابی پایانی در مشاهدات هدف سال ۲۰۲۴ هم‌پوشانی دارند؛ بنابراین، تکرارهای آماری مستقل از یکدیگر محسوب نمی‌شوند، هرچند داده‌های هدف در انتخاب مدل‌ها دخالت نداشته‌اند.

عملکرد پیش‌بینی‌ها در مقیاس اصلی دبی با چهار شاخص ریشه میانگین مربعات خطا[1]، میانگین قدرمطلق خطا[2]، میانگین خطای درصدی مطلق متقارن[3] و ضریب کارایی نش-ساتکلیف[4]، مطابق روابط (۱۲) تا (۱۵)، ارزیابی شد:

جدول ۱ تقسیم زمانی داده‌ها در اعتبارسنجی زمانی و ارزیابی‌های پایانی پیش‌بینی دبی ماهانه.

Table 1 Training and target periods for temporal validation and final evaluation of monthly streamflow forecasts.

Evaluation	Parameter estimation period	12-month target period	24-month target period
Temporal validation: Dec 2018 origin	Jan 1994–Dec 2018	Jan–Dec 2019	Jan 2019–Dec 2020
Temporal validation: Dec 2019 origin	Jan 1994–Dec 2019	Jan–Dec 2020	Jan 2020–Dec 2021
Temporal validation: Dec 2020 origin	Jan 1994–Dec 2020	Jan–Dec 2021	Jan 2021–Dec 2022
Final evaluation: 12 months	Jan 1994–Dec 2023	Jan–Dec 2024	—
Final evaluation: 24 months	Jan 1994–Dec 2022	—	Jan 2023–Dec 2024

[1] Root mean square error (RMSE)
[2] Mean absolute error (MAE)
[3] Symmetric mean absolute percentage error (SMAPE)
[4] Nash–Sutcliffe efficiency (NSE)

$$RMSE=\sqrt{\frac{1}{N}\sum_{t=1}^{N}(Q_{obs,t}-Q_{pred,t})^2} \quad (1)$$

$$MAE=\frac{1}{N}\sum_{t=1}^{N}\left|Q_{obs,t}-Q_{pred,t}\right| \quad (2)$$

$$SMAPE=\frac{100}{N}\sum_{t=1}^{N}\frac{\left|Q_{obs,t}-Q_{pred,t}\right|}{(\left|Q_{obs,t}\right|+\left|Q_{pred,t}\right|)/2} \quad (3)$$

$$NSE=1-\frac{\sum_{t=1}^{N}(Q_{obs,t}-Q_{pred,t})^2}{\sum_{t=1}^{N}(Q_{obs,t}-\bar{Q}_{obs})^2} \quad (4)$$

که در آنها، $Q_{obs,t}$ دبی مشاهده‌شده، $Q_{pred,t}$ دبی پیش‌بینی‌شده مدل در گام زمانی t، $\bar{Q}_{obs}$ میانگین مقادیر مشاهده‌شده، N تعداد مشاهدات دوره آزمون و t شاخص گام زمانی است.

شاخص‌های RMSE و MAE برحسب مترمکعب بر ثانیه، SMAPE به درصد و NSE به‌صورت بدون بعد گزارش شدند. مقادیر کمتر سه شاخص خطا، عملکرد بهتر را نشان می‌دهند. شاخص NSE معرفی‌شده توسط Nash and Sutcliffe (1970)، در حالت انطباق کامل برابر یک است. مقدار صفر بیانگر برابری مجموع مربع خطاها با برآورد ثابتِ میانگین مشاهدات همان دوره و مقدار منفی نشان‌دهنده عملکرد ضعیف‌تر از این مبناست. این مبنا با میانگین ماهانه بلندمدت محاسبه‌شده از داده‌های آموزش تفاوت دارد. شاخص NSE همراه با سایر معیارها تفسیر شد و برای طبقه‌بندی کیفی عملکرد، آستانه ثابتی برای آن در نظر گرفته نشد.

برای بررسی منشأ تفاوت عملکرد دو سناریو برای هر دو مدل مرجع، آمار دبی مشاهده‌شده در دوره‌های هدف و خطاهای ماهانه محاسبه شدند. خطای علامت‌دار برابر اختلاف دبی مشاهده‌شده و پیش‌بینی‌شده تعریف شد؛ بنابراین، مقدار مثبت بیانگر کم‌برآوردی است. افزون بر ارزیابی کل دوره ۲۴ماهه، شاخص‌های عملکرد برای سال‌های ۲۰۲۳ و ۲۰۲۴ جداگانه محاسبه و سهم هر سال از مجموع مربع خطاهای این دوره تعیین شد. پیش‌بینی‌های سال ۲۰۲۴ از دو مبدأ دسامبر ۲۰۲۲ و دسامبر ۲۰۲۳ نیز برای مشاهدات هدف یکسان مقایسه شدند. این مقایسه، تفاوت ترکیب مشاهدات هدف را حذف می‌کند، اما اثر فاصله پیش‌بینی را از تغییر دوره آموزش و پارامترهای برآوردشده به‌طور کامل جدا نمی‌کند.

برای مقایسه عملکرد SARIMA با روش‌های ساده، دو مبنا شامل روش ساده فصلی و میانگین ماهانه بلندمدت به کار رفتند. در روش ساده فصلی، پیش‌بینی هر ماه برابر آخرین مشاهده همان ماه تقویمی در دوره آموزش قرار گرفت؛ بنابراین، برای ارزیابی سال ۲۰۲۴، مقادیر سال ۲۰۲۳ استفاده شدند و در ارزیابی ۲۴ماهه، مقادیر سال ۲۰۲۲ برای هر دو سال هدف تکرار شدند. در روش میانگین ماهانه بلندمدت، میانگین حسابی دبی هر ماه تقویمی در مقیاس اصلی، فقط از داده‌های آموزش متناظر در جدول ۱ محاسبه شد. هیچ‌یک از روش‌های مبنا در طول دوره هدف با مشاهدات جدید به‌روزرسانی نشدند و عملکرد آن‌ها با همان مشاهدات و چهار شاخص مورد استفاده برای SARIMA ارزیابی شد.

۲-۹- ارزیابی ضرورت تفاضل‌گیری غیرفصلی

دو گروه $SARIMA(p,0,q)(P,1,Q)_{12}$ و $SARIMA(p,1,q)(P,1,Q)_{12}$ برای بررسی شواهد مرتبط با افزودن تفاضل‌گیری غیرفصلی مقایسه شدند. در هر دو گروه، مدل‌ها به سری لگاریتمی برازش شدند و تفاضل‌گیری در ساختار مدل اعمال شد. سری‌های حاصل از تفاضل‌گیری صریح، صرفا برای آزمون دیکی-فولر تعمیم‌یافته و بررسی نمودارهای خودهمبستگی و خودهمبستگی جزئی به کار رفتند و دوباره به‌عنوان ورودی مدل‌های دارای تفاضل‌گیری استفاده نشدند.

معیارهای اطلاعاتی آکائیک و بیزی برای غربال ساختارها درون هر گروه، با مرتبه‌های تفاضل‌گیری و داده‌های آموزش یکسان، استفاده شدند. مقایسه دو گروه بر شواهد مانایی، تشخیص باقیمانده‌ها و خطای اعتبارسنجی زمانی پیش از ۲۰۲۳ استوار بود. نتایج ارزیابی‌های پایانی جداگانه گزارش شدند و در محاسبه امتیاز قواعد انتخاب در اعتبارسنجی زمانی وارد نشدند.

این مقایسه برای بررسی سازگاری شواهد تشخیصی و

پیش‌بینی انجام شد. با توجه به تفاوت مرتبه‌های میانگین متحرک در دو مدل مرجع، اختلاف عملکرد آن‌ها به‌تنهایی مبنای تایید یا رد ضرورت تفاضل‌گیری غیرفصلی قرار نگرفت.

۲-۱۰- ارزیابی خطا در ماه‌های کم‌آب و پُرآب

برای بررسی عملکرد هر دو مدل مرجع در شرایط مختلف جریان، ماه‌های دوره ارزیابی بر اساس دبی مشاهده‌شده و آستانه‌های حاصل از دوره آموزش دسته‌بندی شدند. ماه‌های دارای دبی برابر یا کمتر از صدک ۲۵، کم‌آب و ماه‌های دارای دبی برابر یا بیشتر از صدک ۷۵، پُرآب در نظر گرفته شدند؛ سایر ماه‌ها در گروه میانی قرار گرفتند. صدک‌ها با درون‌یابی خطی بر اساس موقعیت‌های احتمال میانی مشاهدات مرتب‌شده، از تمام مقادیر ماهانه دوره آموزش هر سناریو محاسبه شدند، نه به‌صورت جداگانه برای هر ماه تقویمی. دوره‌های آموزش، تعداد مشاهدات و آستانه‌های حاصل در جدول ۲ ارائه شده‌اند. این دسته‌بندی در تحلیل تکمیلی بازنگری برای تفکیک خطا به کار رفت و بیانگر آستانه عملیاتی سیلاب، خشکسالی یا نیاز زیست‌محیطی نیست.

برای کل دوره و گروه‌های کم‌آب و پُرآب، تعداد مشاهدات، ریشه میانگین مربعات خطا، میانگین قدرمطلق خطا، میانگین خطای نسبی علامت‌دار[1] و میانگین قدرمطلق خطای نسبی[2] محاسبه شدند. خطای نسبی علامت‌دار برابر با صد برابر نسبت اختلاف دبی مشاهده‌شده و پیش‌بینی‌شده به دبی مشاهده‌شده تعریف شد؛ بنابراین، مقدار مثبت بیانگر کم‌برآوردی و مقدار منفی بیانگر بیش‌برآوردی است. میانگین قدرمطلق خطای نسبی با نماد MAPE گزارش شد و با شاخص متقارن SMAPE تفاوت دارد. تمام دبی‌های مشاهده‌شده در دوره‌های ارزیابی مثبت بودند.

سوگیری حجمی[3] مطابق رابطه (۱۶) محاسبه شد. برای تبدیل دبی متوسط ماهانه به حجم جریان، دبی هر ماه در مدت همان ماه برحسب ثانیه ضرب شد. سپس اختلاف مجموع حجم مشاهده‌شده و پیش‌بینی‌شده، به‌صورت درصدی از مجموع حجم مشاهده‌شده بیان شد. تعداد واقعی روزهای ماه، شامل فوریه سال کبیسه ۲۰۲۴، در محاسبات لحاظ شد.

$$B_V = 100\frac{\sum_{t\in G}(Q_{obs,t}-Q_{pred,t})\Delta t_t}{\sum_{t\in G}Q_{obs,t}\Delta t_t} \quad (5)$$

در این رابطه، G مجموعه ماه‌های مورد ارزیابی، Δt_t مدت ماه برحسب ثانیه و B_V سوگیری حجمی برحسب درصد است.

۲-۱۱- پیش‌بینی دوره ۲۰۲۵-۲۰۲۶ و تبدیل به مقیاس اصلی

برای تولید پیش‌بینی‌های دبی ماهانه از ژانویه ۲۰۲۵ تا دسامبر ۲۰۲۶، پارامترهای ساختارهای مورد مطالعه با استفاده از کل سری زمانی ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴ دوباره برآورد شدند. پیش‌بینی‌های ۲۴ ماهه از مبدأ دسامبر ۲۰۲۴ و بدون به‌روزرسانی با مشاهدات دوره هدف تولید شدند. پیش‌بینی‌ها ابتدا در مقیاس لگاریتمی به دست آمدند و سپس با تبدیل معکوس رابطه (۱۷) به مقیاس اصلی دبی

جدول ۲ آستانه‌های دسته‌بندی ماه‌های کم‌آب و پُرآب بر اساس صدک‌های دبی دوره آموزش در دو سناریوی ارزیابی.

Table 2 Training-period discharge percentile thresholds for classifying low- and high-flow months in the two evaluation scenarios.

Evaluation period	Training period	Training observations (months)	25th percentile (m^3/s)	75th percentile (m^3/s)
Jan–Dec 2024	Jan 1994–Dec 2023	360	28.612	51.847
Jan 2023–Dec 2024	Jan 1994–Dec 2022	348	28.612	51.997

[1] Mean relative error (MRE)
[2] Mean absolute percentage error (MAPE)
[3] Volume bias (BV)

بازگردانده شدند. همین روش تبدیل معکوس در اعتبارسنجی زمانی و ارزیابی‌های پایانی نیز به کار رفت.

$$\widehat{Q}_{t|T}=\exp\left(\hat{y}_{t|T}\right)\text{-}1 \quad (6)$$

که در آن، T مبدأ پیش‌بینی، ($t>T$) ماه هدف، $\hat{y}_{t|T}$ پیش‌بینی سری لگاریتمی و $\widehat{Q}_{t|T}$ مقدار حاصل از تبدیل معکوس در مقیاس اصلی دبی است. تصحیح سوگیری ناشی از تبدیل معکوس اعمال نشد. تحت فرض نرمال‌بودن توزیع شرطی پیش‌بینی در مقیاس لگاریتمی، رابطه (۱۷) میانه شرطی توزیع پیش‌بینی دبی را به دست می‌دهد و لزوما برابر با میانگین شرطی آن نیست. این تمایز به نوع برآورد نقطه‌ای مربوط است و تعریف متغیر مورد مطالعه، یعنی دبی متوسط ماهانه، را تغییر نمی‌دهد.

در پیاده‌سازی محاسباتی، مقادیر منفی احتمالی حاصل از رابطه (۱۷) به صفر محدود شدند. این اقدام برای رعایت دامنه نامنفی دبی انجام شد و تصحیح سوگیری محسوب نمی‌شود. شاخص‌های عملکرد پیش‌بینی‌های نقطه‌ای نیز با همین تبدیل معکوس و محدودسازی محاسبه شدند. پیش‌بینی‌های دوره ۲۰۲۶-۲۰۲۵، برون‌یابی مدل‌ها از مبدأ دسامبر ۲۰۲۴ هستند و در ارزیابی عملکرد گزارش‌شده این پژوهش نقشی ندارند.

۲-۱۲- بیان عدم‌قطعیت پیش‌بینی

با فرض تقریب نرمال توزیع خطای پیش‌بینی در مقیاس لگاریتمی، بازه‌های پیش‌بینی نقطه‌ای با سطح اسمی ۹۵ درصد بر اساس واریانس خطای پیش‌بینی محاسبه شدند (Box et al., 2015). حدود پایین و بالای بازه در ماه هدف t، از مبدأ پیش‌بینی T، به‌صورت زیر تعریف شدند:

$$L_{t|T}=\hat{y}_{t|T}\text{-}1.96\sqrt{\mathrm{MSE}_{t|T}} \quad (7)$$

$$U_{t|T}=\hat{y}_{t|T}+1.96\sqrt{\mathrm{MSE}_{t|T}} \quad (8)$$

که در آنها، MSE_t واریانس خطای پیش‌بینی در گام زمانی t در مقیاس لگاریتمی، L_t و U_t به‌ترتیب حدود پایین و بالای بازه در این مقیاس هستند. سپس حدود پایین و بالا با استفاده از تبدیل معکوس به مقیاس اصلی دبی منتقل شدند:

$$Q_{L,t|T}=\exp\left(L_{t|T}\right)\text{-}1 \quad (9)$$

$$Q_{U,t|T}=\exp\left(U_{t|T}\right)\text{-}1 \quad (10)$$

که در آنها، $Q_{L,t|T}$ و $Q_{U,t|T}$ به‌ترتیب حدود پایین و بالای بازه پیش‌بینی دبی هستند. تبدیل معکوس به هر دو حد اعمال شد؛ بنابراین، بازه حاصل در مقیاس اصلی الزاما نسبت به پیش‌بینی نقطه‌ای متقارن نیست. در پیاده‌سازی، حدود پایین منفی احتمالی به صفر محدود شدند.

پوشش بازه‌ها برای هر دو مدل مرجع در دوره‌های ارزیابی ۱۲ و ۲۴ماهه بررسی شد. در هر سناریو، پیش‌بینی و میانگین مربعات خطای متناظر از مدل برازش‌شده به داده‌های آموزش همان سناریو به دست آمدند. پوشش تجربی برابر درصد مشاهداتی تعریف شد که بین حدود پایین و بالای بازه، با احتساب خود حدود، قرار گرفتند. میانگین پهنای بازه در مقیاس اصلی دبی و تعداد مشاهدات پایین‌تر از حد پایین و بالاتر از حد بالا نیز محاسبه شدند.

این بازه‌ها برای مقادیر آینده دبی تعریف شده‌اند و با بازه اطمینان پارامترهای مدل تفاوت دارند. سطح اسمی ۹۵ درصد به هر گام پیش‌بینی مربوط است و پوشش هم‌زمان کل مسیر را تضمین نمی‌کند. بازه‌ها مشروط به ساختار و پارامترهای برآوردشده مدل‌اند و عدم‌قطعیت ناشی از برآورد پارامترها و انتخاب ساختار به‌صورت جداگانه در آن‌ها لحاظ نشده است.

۲-۱۳- محیط محاسباتی

برازش مدل‌های ARIMA و SARIMA، آزمون‌های آماری، اعتبارسنجی زمانی و تولید پیش‌بینی‌ها در محیط متلب و با استفاده از جعبه‌ابزار اقتصادسنجی انجام شدند. برای برآورد پارامترها، استخراج باقیمانده‌ها و محاسبه پیش‌بینی‌ها و میانگین مربعات خطای پیش‌بینی، به‌ترتیب از توابع estimate،infer و forecast استفاده شد. محاسبه

معیارهای اطلاعاتی، امتیازهای اعتبارسنجی و شاخص‌های عملکرد نیز در کدهای تحلیل پیاده‌سازی شد.

۳- نتایج و بحث

۳-۱- تحلیل توصیفی و بررسی ویژگی‌های اولیه سری زمانی دبی

سری دبی متوسط ماهانه رودخانه سالمون شامل ۳۷۲ مشاهده از ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴ بررسی شد. مطابق شکل ۱، افزایش و کاهش دبی در طول دوره الگوی تکرارشونده‌ای دارد، اما بزرگی قله‌های ماهانه میان سال‌ها متفاوت است. میانگین، انحراف معیار، حداقل و حداکثر دبی به‌ترتیب برابر با ۵۲/۶۲، ۴۷/۶۱، ۱۱/۳۹ و ۳۳۲/۵۵ مترمکعب بر ثانیه بود. این مقادیر و تغییرات نمایش‌داده‌شده در شکل ۱، دامنه گسترده نوسانات دبی ماهانه و تفاوت بزرگی جریان‌ها میان سال‌ها را نشان می‌دهند.

نتایج آزمون دیکی-فولر تعمیم‌یافته برای سری لگاریتمی به مشخصات رگرسیون آزمون وابسته بود (جدول ۳). در حالت بدون ثابت و روند و با صفر وقفه افزوده، فرض صفر وجود ریشه واحد رد نشد؛ مقدار احتمال برابر با ۰/۲۱۷۶ بود. با لحاظ‌کردن ثابت و بدون روند خطی، فرض صفر در حالت‌های صفر و ۱۲ وقفه افزوده رد شد، اما در حالت ۲۴ وقفه افزوده رد نشد. بنابراین، نتیجه اجرای اولیه به‌تنهایی برای نامانا دانستن قطعی سری لگاریتمی کافی نبود.

برای هر دو سری دارای فقط تفاضل‌گیری فصلی و دارای تفاضل‌گیری معمولی و فصلی، فرض صفر در تمام مشخصات بررسی‌شده رد شد. بااین‌حال، این نتیجه ضرورت افزودن تفاضل‌گیری معمولی را تایید نمی‌کند؛ زیرا در حالت دارای فقط تفاضل فصلی نیز فرض صفر رد شده است. آماره منفی‌تر در حالت تفاضل ترکیبی نیز به‌تنهایی نشان‌دهنده برتری آن برای پیش‌بینی نیست. افزون بر این، آزمون دیکی-فولر تعمیم‌یافته به‌تنهایی ضرورت تفاضل‌گیری فصلی را تعیین نمی‌کند؛ ازاین‌رو، نتایج آن همراه با نمودارهای همبستگی و ارزیابی پیش‌بینی تفسیر شدند.

در جدول ۳، مشخصات AR فاقد ثابت و روند خطی و مشخصات ARD شامل ثابت و فاقد روند خطی است. تعداد وقفه‌های افزوده به‌صورت ثابت تعیین شد و انتخاب خودکار وقفه انجام نشد. سری لگاریتمی بدون تفاضل شامل ۳۷۲ مشاهده از ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴، سری دارای فقط تفاضل فصلی شامل ۳۶۰ مشاهده از ژانویه ۱۹۹۵ تا دسامبر ۲۰۲۴ و سری دارای تفاضل معمولی و فصلی شامل ۳۵۹ مشاهده از فوریه ۱۹۹۵ تا دسامبر ۲۰۲۴ بود. تعداد مشاهدات قابل استفاده در رگرسیون آزمون، به‌دلیل تفاضل‌گیری در معادله آزمون و لحاظ‌کردن وقفه‌های افزوده، از طول سری ورودی کمتر است. تصمیم‌های رد و عدم رد به فرض صفر وجود ریشه واحد اشاره دارند. مقادیر احتمال ۰/۰۰۱ که با علامت ستاره مشخص شده‌اند، حد پایین گزارش تابع adftest هستند و نباید مقدار احتمال دقیق تلقی شوند.

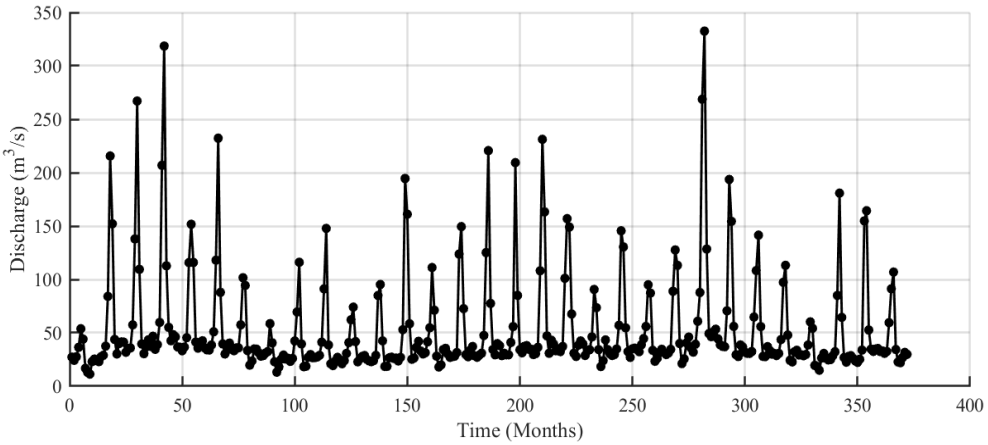


Fig. 1 Monthly discharge time series of the Salmon River from January 1994 to December 2024.

شکل ۱ تغییرات سری زمانی دبی ماهانه رودخانه سالمون در دوره ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴.

جدول ۳ نتایج آزمون دیکی-فولر تعمیم‌یافته برای سری لگاریتمی و دو حالت تفاضل‌گیری، با مشخصات متفاوت جزء ثابت و تعداد وقفه‌های افزوده.

Table 3 Augmented Dickey–Fuller test results for the log-transformed series and two differencing schemes under alternative intercept and augmentation-lag specifications.

Series	Specification	Augmentation lags	Regression observations	ADF statistic	5% critical value	p-value	Decision at 5%
Log-transformed, undifferenced	AR	0	371	−1.1847	−1.9415	0.2176	Do not reject H_0
Log-transformed, undifferenced	ARD	0	371	−8.2682	−2.8698	0.001*	Reject H_0
Log-transformed, undifferenced	ARD	12	359	−3.5050	−2.8700	0.0088	Reject H_0
Log-transformed, undifferenced	ARD	24	347	−2.3315	−2.8701	0.1631	Do not reject H_0
d=0, D=1	AR	0	359	−7.7468	−1.9415	0.001*	Reject H_0
d=0, D=1	ARD	0	359	−7.7427	−2.8700	0.001*	Reject H_0
d=0, D=1	ARD	12	347	−5.6999	−2.8701	0.001*	Reject H_0
d=0, D=1	ARD	24	335	−3.8877	−2.8703	0.0031	Reject H_0
d=1, D=1	AR	0	358	−20.3403	−1.9415	0.001*	Reject H_0
d=1, D=1	ARD	0	358	−20.3118	−2.8700	0.001*	Reject H_0
d=1, D=1	ARD	12	346	−7.5257	−2.8702	0.001*	Reject H_0
d=1, D=1	ARD	24	334	−5.9785	−2.8703	0.001*	Reject H_0

۳-۲- بررسی ساختار خودهمبستگی سری لگاریتمی

توابع خودهمبستگی و خودهمبستگی جزئی سری لگاریتمی دبی ماهانه تا وقفه ۴۰ ماه در شکل ۲ ارائه شده‌اند. در تابع خودهمبستگی، ضریب وقفه ۱ مثبت و خارج از حدود مرجع است. قله‌های مثبت در وقفه‌های ۱۲، ۲۴ و ۳۶ ماه و همبستگی‌های منفی در تعدادی از وقفه‌های بین آن‌ها، الگوی تناوبی سالانه را نشان می‌دهند. دامنه قله‌های سالانه با افزایش وقفه کاهش می‌یابد، اما هر سه قله همچنان از حد بالای مرجع فراتر هستند.

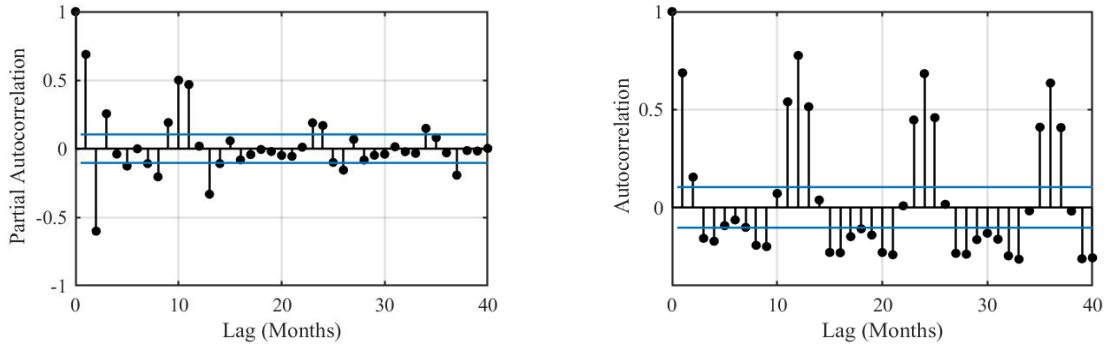
در تابع خودهمبستگی جزئی، ضریب وقفه ۱ مثبت و ضریب وقفه ۲ منفی است و هر دو از حدود مرجع عبور می‌کنند. در برخی وقفه‌های نزدیک به دوره سالانه و مضرب‌های آن نیز ضرایب خارج از این حدود دیده می‌شوند؛ بااین‌حال، الگو به قطع‌شدن مشخص ضرایب پس از یک وقفه معین محدود نیست. ازاین‌رو، این نمودارها برای شناسایی وابستگی‌های کوتاه‌مدت و سالانه استفاده شدند و تعیین مرتبه‌های مدل صرفا بر مبنای آن‌ها انجام نشد. وجود قله‌های سالانه در تابع خودهمبستگی، به‌تنهایی ضرورت تفاضل‌گیری فصلی یا افزودن تفاضل‌گیری معمولی را مشخص نمی‌کند. برای بررسی اثر تبدیل‌ها بر ساختار وابستگی، نمودارهای دو حالت دارای فقط تفاضل فصلی و دارای تفاضل معمولی و فصلی نیز بررسی شدند.

۳-۳- بررسی ساختار خودهمبستگی پس از تفاضل‌گیری

توابع خودهمبستگی و خودهمبستگی جزئی سری لگاریتمی پس از یک مرتبه تفاضل‌گیری فصلی با دوره ۱۲ ماه، و پس از ترکیب آن با یک مرتبه تفاضل‌گیری معمولی، در شکل ۳ مقایسه شده‌اند. در حالت دارای فقط تفاضل فصلی، خودهمبستگی مثبت در وقفه‌های ابتدایی باقی مانده و دامنه آن با افزایش وقفه کاهش یافته است. در حوالی وقفه ۱۲ نیز ضرایب منفی خارج از حدود مرجع دیده می‌شوند (شکل ۳،آ). نمودار خودهمبستگی جزئی، علاوه بر ضریب مثبت بزرگ در وقفه نخست، ضرایب خارج از حدود مرجع را در حوالی وقفه‌های ۱۲ و ۲۴ نشان می‌دهد (شکل ۳،ب). بنابراین، سری پس از تفاضل فصلی همچنان دارای وابستگی زمانی است.

با افزودن تفاضل‌گیری معمولی، الگوی خودهمبستگی در وقفه‌های ابتدایی تغییر کرده و چند ضریب منفی ظاهر شده است. در این حالت، خودهمبستگی منفی برجسته در وقفه ۱۲ و ضرایب خارج از حدود مرجع در برخی وقفه‌های دیگر

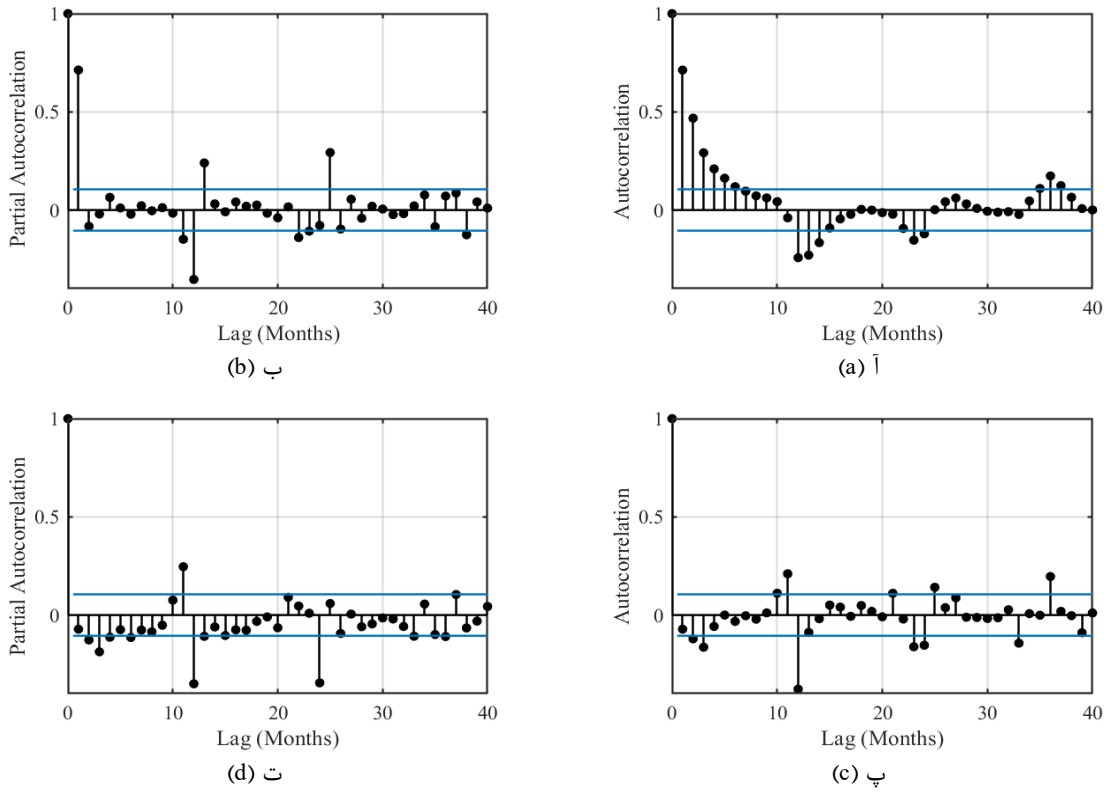
باقی مانده‌اند (شکل ۳،پ). نمودار خودهمبستگی جزئی نیز ضرایب منفی مشخصی در وقفه‌های ۱۲ و ۲۴ نشان می‌دهد (شکل ۳، ت).



آ(a) ب(b)

Fig. 2 (a) Autocorrelation function and (b) partial autocorrelation function of the log-transformed monthly mean discharge series of the Salmon River before differencing, up to a lag of 40 months. Horizontal lines on either side of zero indicate approximate 95% reference bounds.

شکل ۲ (آ) تابع خودهمبستگی و (ب) تابع خودهمبستگی جزئی سری لگاریتمی دبی متوسط ماهانه رودخانه سالمون، پیش از تفاضل‌گیری، تا وقفه ۴۰ ماه. خطوط افقی دو سوی صفر، حدود مرجع تقریبی ۹۵ درصد را نشان می‌دهند.



(a) آ (b) ب

(c) پ (d) ت

Fig. 3 Autocorrelation and partial autocorrelation functions of the log-transformed series after differencing: (a) ACF and (b) PACF for d=0 and D=1; (c) ACF and (d) PACF for d=1 and D=1.

شکل ۳ توابع خودهمبستگی و خودهمبستگی جزئی سری لگاریتمی پس از تفاضل‌گیری: (آ) خودهمبستگی و (ب) خودهمبستگی جزئی در حالت d=0 و D=1 ؛ (پ) خودهمبستگی و (ت) خودهمبستگی جزئی در حالت d=1 و D=1.

این تفاوت‌ها نشان می‌دهند که افزودن تفاضل‌گیری معمولی، ساختار وابستگی سری را تغییر داده است؛ بااین‌حال، کاهش خودهمبستگی مثبت در وقفه‌های ابتدایی به‌تنهایی ضرورت این تبدیل را تایید نمی‌کند. ضرایب منفی مشاهده‌شده نیز برای تشخیص قطعی تفاضل‌گیری بیش‌ازحد کافی نیستند. ازاین‌رو، نمودارها همراه با نتایج آزمون ریشه واحد، تشخیص باقیمانده‌ها و اعتبارسنجی زمانی بررسی شدند و مبنای مستقل ترجیح یکی از دو گروه تفاضل‌گیری قرار نگرفتند. در شکل ۳، دوره فصلی ۱۲ ماه است و نمودارها تا وقفه ۴۰ ماه نمایش داده شده‌اند. خطوط افقی دو سوی صفر، حدود مرجع تقریبی ۹۵ درصد را نشان می‌دهند.

۳-۴- نتایج اعتبارسنجی زمانی قواعد انتخاب مبتنی بر معیارهای اطلاعاتی

ساختارهای پیشنهادی معیارهای اطلاعاتی آکائیک و بیزی در سه مبدأ اعتبارسنجی در جدول ۴ ارائه شده‌اند. در هر مبدأ، معیارها از برازش ساختارها به داده‌های ژانویه ۱۹۹۴ تا دسامبر همان سال محاسبه شدند و مقایسه آن‌ها به مدل‌های دارای دوره آموزش و مرتبه‌های تفاضل‌گیری یکسان محدود شد.

در گروه دارای فقط تفاضل‌گیری فصلی، هر دو معیار در هر سه مبدأ ساختار $SARIMA(1,0,0)(1,1,2)_{12}$ را پیشنهاد کردند. در گروه دارای تفاضل‌گیری معمولی و فصلی، پیشنهاد دو معیار متفاوت بود: آکائیک در هر سه مبدأ ساختار $SARIMA(2,1,2)(1,1,2)_{12}$ و بیزی ساختار $SARIMA(1,1,2)(1,1,2)_{12}$ را انتخاب کرد.

عملکرد پیش‌بینی قواعد انتخاب در جدول ۵ خلاصه شده است. در گروه دارای فقط تفاضل فصلی، یکسان‌بودن ساختارهای پیشنهادی به برابری خطاهای اعتبارسنجی دو قاعده منجر شد. بر اساس قاعده برخورد با تساوی، بیزی ترجیح داده شد؛ این ترجیح در این گروه تغییری در ساختار یا پیش‌بینی‌ها ایجاد نکرد. در گروه دارای تفاضل ترکیبی، قاعده بیزی در هر دو افق، میانگین RMSE کمتری داشت و بر اساس امتیاز تجمیعی ترجیح داده شد.

میانگین‌های هر افق در جدول ۵ از سه مبدأ اعتبارسنجی محاسبه شدند و امتیاز انتخاب، میانگین حسابی شش مقدار

جدول ۴ ساختارهای پیشنهادی معیارهای اطلاعاتی آکائیک و بیزی در هر گروه تفاضل‌گیری و مبدأ اعتبارسنجی.

Table 4 SARIMA structures selected by AIC and BIC within each differencing group at the temporal validation origins.

Differencing group	Forecast origin	Selection rule	Selected SARIMA structure	Training AIC	Training BIC
d=0, D=1	Dec 2018	AIC / BIC	$(1,0,0)(1,1,2)_{12}$	−84.628	−66.109
d=0, D=1	Dec 2019	AIC / BIC	$(1,0,0)(1,1,2)_{12}$	−98.156	−79.441
d=0, D=1	Dec 2020	AIC / BIC	$(1,0,0)(1,1,2)_{12}$	−113.252	−94.349
d=1, D=1	Dec 2018	AIC	$(2,1,2)(1,1,2)_{12}$	−84.045	−54.415
d=1, D=1	Dec 2018	BIC	$(1,1,2)(1,1,2)_{12}$	−81.796	−55.870
d=1, D=1	Dec 2019	AIC	$(2,1,2)(1,1,2)_{12}$	−97.457	−67.513
d=1, D=1	Dec 2019	BIC	$(1,1,2)(1,1,2)_{12}$	−94.698	−68.497
d=1, D=1	Dec 2020	AIC	$(2,1,2)(1,1,2)_{12}$	−112.318	−82.072
d=1, D=1	Dec 2020	BIC	$(1,1,2)(1,1,2)_{12}$	−109.713	−83.247

جدول ۵ مقایسه عملکرد قواعد انتخاب مبتنی بر آکائیک و بیزی در سه مبدأ اعتبارسنجی و دو افق پیش‌بینی.

Table 5 Forecast performance of AIC- and BIC-based selection rules across three temporal validation origins and two forecast horizons.

Differencing group	Selection rule	Mean RMSE, 12 months (m^3/s)	Mean RMSE, 24 months (m^3/s)	Selection score (m^3/s)	Rule decision
d=0, D=1	AIC	15.305	19.552	17.428	Tied
d=0, D=1	BIC	15.305	19.552	17.428	Preferred by tie-break
d=1, D=1	AIC	17.881	22.786	20.333	Not preferred
d=1, D=1	BIC	16.078	21.659	18.868	Preferred by lower score

RMSE بود. ترجیح بیزی در گروه تفاضل ترکیبی به معنای خطای کمتر در تمام مبدأها نبود؛ در مبدأ دسامبر ۲۰۱۸، ساختار پیشنهادی آکائیک در هر دو افق خطای کمتری داشت، اما در مبدأهای دسامبر ۲۰۱۹ و ۲۰۲۰، خطای ساختار پیشنهادی بیزی کمتر بود. بنابراین، نتیجه انتخاب بر عملکرد تجمیعی متکی بود

پس از تعیین قاعده ترجیحی در گروه تفاضل ترکیبی، اعمال معیار بیزی بر ساختارهای برازش‌شده به ۳۴۸ مشاهده ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۲، به انتخاب $SARIMA(1,1,2)(1,1,2)_{12}$ انجامید. مقادیر آکائیک و بیزی این برازش به‌ترتیب ۱۱۳/۲۲۰- و ۸۶/۲۵۵- بودند. ساختار حاصل در ارزیابی‌های پایانی ثابت نگه داشته شد. در گروه دارای فقط تفاضل فصلی، مدل مرجع $SARIMA(1,0,0)(1,1,2)_{12}$ نیز با پیشنهاد هر دو معیار در تمام مبدأهای اعتبارسنجی یکسان بود؛ بااین‌حال، در این گروه مرحله جداگانه انتخاب مجدد ساختار بر داده‌های تا دسامبر ۲۰۲۲ اجرا نشد.

در مقایسه دو گروه، مدل دارای فقط تفاضل فصلی، میانگین RMSE کمتری در هر دو افق و امتیاز تجمیعی پایین‌تری داشت. این نتیجه از عملکرد بهتر آن بر اساس معیار تجمیعی اعتبارسنجی حاضر پشتیبانی می‌کند، اما به معنای برتری در همه مبدأها یا دوره‌های دیگر نیست. هر دو ساختار مرجع برای بررسی تفاوت نتایج تشخیصی و عملکرد پیش‌بینی در ادامه تحلیل حفظ شدند.

۳-۵- مقایسه پیچیدگی مدل‌های مرجع با ساختارهای ساده‌تر

برای بررسی ارتباط پیچیدگی ساختار با عملکرد پیش‌بینی، مدل‌های مرجع با ساختارهای ساده‌تر در همان گروه تفاضل‌گیری مقایسه شدند. در گروه دارای فقط تفاضل فصلی، ۱۲ ساختار و در گروه دارای تفاضل معمولی و فصلی، ۳۶ ساختار، با احتساب مدل مرجع، بررسی شدند. هر ساختار در سه مبدأ دسامبر ۲۰۱۸، ۲۰۱۹ و ۲۰۲۰ برازش شد. هر ۳۶ برازش گروه نخست و هر ۱۰۸ برازش گروه دوم موفق بودند. نتایج مدل‌های مرجع و تعدادی از ساختارهای ساده‌تر در جدول ۶ ارائه شده‌اند.

در گروه دارای فقط تفاضل فصلی، مدل $SARIMA(1,0,0)(1,1,2)_{12}$ کمترین امتیاز تجمیعی را در مجموعه ۱۲ساختاری داشت. امتیاز آن نسبت به کم‌خطاترین ساختار ساده‌تر، $SARIMA(1,0,0)(0,1,2)_{12}$، به میزان ۶/۶۷ درصد کمتر بود. بااین‌حال، فرض صفر آزمون لیونگ-باکس تا وقفه ۱۲ در هر سه مبدأ برای مدل مرجع رد شد؛ در آزمون تا وقفه ۲۴، فرض صفر در هیچ‌یک از این مبدأها رد نشد.

در گروه دارای تفاضل ترکیبی، ۳۰ ساختار ساده‌تر امتیاز تجمیعی کمتری از مدل مرجع $SARIMA(1,1,2)(1,1,2)_{12}$ داشتند. کمترین امتیاز به $SARIMA(0,1,1)(1,1,2)_{12}$ تعلق داشت و نسبت به مدل مرجع ۲۲/۷۵ درصد کمتر بود. در مقابل، مدل مرجع در هر سه مبدأ کمترین مقدار آکائیک و بیزی را در همین مجموعه ۳۶ ساختاری داشت و تنها ساختاری بود که فرض صفر آزمون لیونگ-باکس در هر دو وقفه ۱۲ و ۲۴، در هر سه مبدأ، برای آن رد نشد.

جدول ۶ مقایسه مدل‌های مرجع و ساختارهای ساده‌تر منتخب در اعتبارسنجی زمانی با مبدأهای دسامبر ۲۰۱۸، ۲۰۱۹ و ۲۰۲۰.

Table 6 Comparison of the reference models and selected simpler structures in temporal validation at the December 2018, 2019, and 2020 forecast origins.

SARIMA structure	K	Mean RMSE, 12 months (m^3/s)	Mean RMSE, 24 months (m^3/s)	Aggregate score (m^3/s)	LB rejections: lag 12	LB rejections: lag 24
$(1,0,0)(1,1,2)_{12}$	5	15.305	19.552	17.428	3	0
$(1,0,0)(0,1,2)_{12}$	4	16.463	20.883	18.673	3	0
$(1,1,2)(1,1,2)_{12}$	7	16.078	21.659	18.868	0	0
$(0,1,1)(1,1,2)_{12}$	5	12.458	16.694	14.576	3	3
$(1,1,0)(1,1,2)_{12}$	5	12.472	16.707	14.589	3	3
$(0,1,0)(1,1,2)_{12}$	4	12.484	16.714	14.599	3	3

در جدول ۶، تعداد پارامترهای آزاد با K مشخص شده و شامل ضرایب خودرگرسیونی و میانگین متحرک و واریانس نوآوری است؛ ثابتِ تثبیت‌شده در صفر در این شمارش منظور نشده است. امتیاز تجمیعی، میانگین شش مقدار RMSE حاصل از سه مبدأ و دو افق است. دو ستون پایانی، تعداد رد فرض صفر آزمون لیونگ-باکس را در سطح معنی‌داری ۵ درصد، از مجموع سه برازش آموزشی نشان می‌دهند.

این مقایسه نشان داد که افزایش پیچیدگی در دو گروه پیامد یکسانی نداشت. در گروه دارای فقط تفاضل فصلی، ساختار مرجع نسبت به ساختارهای ساده‌تر بررسی‌شده خطای تجمیعی کمتری داشت، هرچند محدودیت تشخیصی باقیمانده‌های آن باقی بود. در گروه تفاضل ترکیبی، مقادیر کمتر معیارهای اطلاعاتی و عدم رد آزمون‌های باقیمانده‌ها برای مدل مرجع، با کمترین خطای اعتبارسنجی همراه نشد. بنابراین، پیچیدگی بیشتر این مدلاز منظر کاهش خطای پیش‌بینی در مجموعه بررسی‌شده توجیه نشد.

ساختارهای ساده‌تر بر اساس آزمون باقیمانده‌ها از مقایسه کنار گذاشته نشدند؛ ازاین‌رو، خطای کمتر آن‌ها باید در تفسیر نتایج حفظ شود. دو مدل مرجع برای ادامه مقایسه نگه داشته شدند، اما مدل دارای تفاضل ترکیبی به‌عنوان کم‌خطاترین ساختار معرفی نمی‌شود.

۳-۶- برآورد ضرایب دو مدل مرجع

ضرایب دو مدل $SARIMA(1,0,0)(1,1,2)_{12}$ و $SARIMA(1,1,2)(1,1,2)_{12}$ با برازش مجدد به کل سری ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴ برآورد شدند. برآورد پارامترها و خطاهای استاندارد متناظر در جدول ۷ ارائه شده‌اند. این مقادیر به برازش کل دوره مربوط‌اند و با برآوردهای جداگانه هر مبدأ اعتبارسنجی و هر سناریوی ارزیابی پایانی تفاوت دارند. ثابت هر دو مدل در صفر تثبیت شد و برای آن خطای استاندارد گزارش نشد.

اعداد داخل پرانتز در ستون پارامتر، وقفه را برحسب ماه نشان می‌دهند. خط تیره برای ضرایب میانگین متحرک مدل نخست، بیانگر نبود آن جمله در ساختار مدل است؛ در ردیف ثابت نیز به گزارش‌نشدن خطای استاندارد برای پارامتر تثبیت‌شده اشاره دارد.

در هر دو مدل، ضریب خودرگرسیونی غیرفصلی مثبت و ضرایب خودرگرسیونی فصلی و میانگین متحرک فصلی منفی برآورد شدند. ضریب میانگین متحرک فصلی در وقفه ۱۲ در هیچ‌یک از دو مدل در سطح ۵ درصد معنادار نبود؛ مقادیر احتمال متناظر برای مدل‌های دارای $d=0$ و $d=1$ به‌ترتیب ۰/۲۵۵۳ و ۰/۱۷۱۹ بودند. سایر ضرایب پویای برآوردشده در این سطح معنادار بودند. بنابراین، نتایج از معناداری تک‌تک جمله‌های دو مدل پشتیبانی نمی‌کنند.

مقایسه ساختارهای دارای Q=1 و Q=2 در تحلیل پیچیدگی، حذف جمله میانگین متحرک فصلی در وقفه ۲۴ را بررسی می‌کند و معادل حذف اختصاصی جمله وقفه ۱۲ نیست؛ در نتیجه، آن تحلیل به‌تنهایی درباره ضرورت حفظ ضریب نامعنادار وقفه ۱۲ پاسخ نمی‌دهد. واریانس نوآوری‌ها نیز در مقیاس لگاریتمی گزارش شده است و به‌تنهایی معیار دقت پیش‌بینی در مقیاس اصلی دبی محسوب نمی‌شود.

۳-۷- تحلیل باقیمانده‌ها و مقایسه کفایت تشخیصی دو مدل مرجع

هر ۱۸ برازش آریما معیارهای پذیرش عددی را برآورده کردند. در گروه بدون تفاضل‌گیری غیرفصلی، معیارهای اطلاعاتی ساختار یکسانی را پیشنهاد نکردند: کمترین AIC به ARIMA(2,0,2) و کمترین BIC به ARIMA(1,0,1) تعلق داشت. در گروه دارای یک مرتبه تفاضل‌گیری غیرفصلی، ساختار ARIMA(2,1,2) کمترین مقدار هر دو معیار را داشت. برای هر سه ساختار، فرض صفر نبود خودهمبستگی باقیمانده‌ها در آزمون‌های لیونگ-باکس تا وقفه‌های ۱۲ و ۲۴ رد شد (جدول ۸).

بنابراین، ساختارهای پیشنهادی معیارهای اطلاعاتی در مجموعه غیرفصلی بررسی‌شده، وابستگی زمانی سری را به‌طور کامل بازنمایی نکردند. در جدول ۸، Kتعداد پارامترهای آزاد، شامل ضرایب خودرگرسیونی و میانگین متحرک و واریانس نوآوری است؛ ثابت مدل در صفر تثبیت

شد و در این شمارش منظور نشد. ترجیح معیارهای اطلاعاتی فقط درون هر گروه d تعیین شد و مقادیر دو گروه برای رتبه‌بندی مشترک استفاده نشدند. درجه آزادی آزمون لیونگ-باکس برابر با $df=m-p-q$ است که در آن m حداکثر وقفه آزمون است. فرض صفر، نبود خودهمبستگی در وقفه‌های ۱ تا m است و در تمام موارد جدول ۸، در سطح معنی‌داری ۵ درصد رد شد. مقادیر احتمال که در خروجی نرم‌افزار صفر ثبت شده‌اند، به‌صورت $p<0.001$ گزارش شدند و به معنای احتمال دقیقا صفر نیستند.

باقیمانده‌های حاصل از برازش دو مدل مرجع به کل سری ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴ در مقیاس لگاریتمی بررسی شدند. نتایج آزمون لیونگ-باکس با درجه آزادی تعدیل‌شده در جدول ۹ ارائه شده‌اند. این نتایج به برازش کل دوره مربوط‌اند و از آزمون‌های باقیمانده‌های آموزشی در مبدأهای اعتبارسنجی، گزارش‌شده در بخش مقایسه پیچیدگی، متمایزند.

برای مدل دارای فقط تفاضل فصلی، فرض صفر نبود خودهمبستگی تا وقفه ۱۲ در سطح معنی‌داری ۵ درصد رد شد، اما در آزمون تا وقفه ۲۴ رد نشد. برای مدل دارای تفاضل معمولی و فصلی، فرض صفر در هیچ‌یک از دو آزمون رد نشد. این نتایج بیانگر تفاوت تشخیصی دو ساختارند؛ بااین‌حال، عدم رد فرض صفر به معنای اثبات استقلال باقیمانده‌ها یا برتری پیش‌بینی مدل نیست.

جدول ۷ برآورد پارامترها و خطاهای استاندارد دو مدل مرجع ساریما بر اساس برازش به داده‌های ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴.

Table 7 Parameter estimates and standard errors of the two reference SARIMA models fitted to data from January 1994 to December 2024.

Parameter	Estimate: $(1,0,0)(1,1,2)_{12}$	Standard error	Estimate: $(1,1,2)(1,1,2)_{12}$	Standard error
Constant	0 (fixed)	—	0 (fixed)	—
AR(1)	0.77307	0.02847	0.52539	0.07822
SAR(12)	−0.74089	0.02765	−0.75135	0.02629
MA(1)	—	—	−0.71909	0.09379
MA(2)	—	—	−0.17062	0.07709
SMA(12)	−0.04993	0.04389	−0.06141	0.04495
SMA(24)	−0.75839	0.03618	−0.76643	0.03701
Innovation variance	0.040616	0.001894	0.040613	0.001910

جدول ۸ معیارهای اطلاعاتی و نتایج آزمون لیونگ-باکس برای ساختارهای آریما پیشنهادی در هر گروه تفاضل‌گیری، برازش‌شده به کل دوره ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴.

Table 8 Information criteria and Ljung–Box test results for the ARIMA structures preferred within each differencing group, fitted to the full January 1994–December 2024 series.

ARIMA structure	Preferred by	K	AIC	BIC	Maximum lag	df	Q statistic	p-value
(2,0,2)	AIC	5	390.208	409.802	12	8	262.396	<0.001
(2,0,2)	AIC	5	390.208	409.802	24	20	507.943	<0.001
(1,0,1)	BIC	3	394.464	406.221	12	10	310.543	<0.001
(1,0,1)	BIC	3	394.464	406.221	24	22	601.327	<0.001
(2,1,2)	AIC and BIC	5	257.761	277.356	12	8	157.671	<0.001
(2,1,2)	AIC and BIC	5	257.761	277.356	24	20	310.723	<0.001

جدول ۹ نتایج آزمون لیونگ-باکس با درجه آزادی تعدیل‌شده برای باقیمانده‌های دو مدل مرجع برازش‌شده به کل دوره ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴.

Table 9 Ljung–Box test results with adjusted degrees of freedom for residuals of the two reference models fitted to the full January 1994–December 2024 series.

SARIMA structure	Maximum lag (months)	Q statistic	Degrees of freedom	p-value	Decision at 5%
$(1,0,0)(1,1,2)_{12}$	12	17.028	8	0.030	Reject H_0
$(1,0,0)(1,1,2)_{12}$	24	30.303	20	0.065	Do not reject H_0
$(1,1,2)(1,1,2)_{12}$	12	7.915	6	0.244	Do not reject H_0
$(1,1,2)(1,1,2)_{12}$	24	21.614	18	0.250	Do not reject H_0

فرض صفر آزمون‌های جدول ۹، نبود خودهمبستگی در وقفه‌های ۱ تا حداکثر وقفه آزمون است. بنابراین، رد آزمون تا وقفه ۱۲ به‌تنهایی نشان نمی‌دهد که خودهمبستگی باقی‌مانده مشخصا به وقفه ۱۲ یا مولفه فصلی مربوط است.

نمودارهای تشخیصی دو مدل در شکل ۴ مقایسه شده‌اند. در هر دو مدل، باقیمانده‌ها حول صفر نوسان دارند و چند مقدار با دامنه بزرگ‌تر مشاهده می‌شود، اما روند یا الگوی تناوبی آشکاری در نمودارهای زمانی دیده نمی‌شود (شکل ۴، آ و ب). در نمودارهای خودهمبستگی جزئی تا وقفه ۲۰، بیشتر ضرایب داخل حدود مرجع قرار دارند؛ بااین‌حال، ضرایبی نزدیک به این حدود یا خارج از آن‌ها، به‌ویژه در حوالی وقفه ۱۱، مشاهده می‌شوند (شکل ۴،پ و ت). ارزیابی این نمودارها همراه با نتایج آزمون لیونگ-باکس انجام شد و مبنای تایید استقلال یا ثبات واریانس باقیمانده‌ها قرار نگرفت.

در نمودارهای چندک-چندک هر دو مدل، نقاط بخش مرکزی به خط مرجع نزدیک‌ترند، اما در دو انتهای توزیع انحراف آشکاری دیده می‌شود (شکل ۴، ث و ج).

بنابراین، انطباق با توزیع نرمال در سراسر دامنه باقیمانده‌ها مشاهده نمی‌شود. این یافته نشان می‌دهد که نتایج آزمون خودهمبستگی و بررسی شکل توزیع، جنبه‌های متفاوتی از کفایت مدل را توصیف می‌کنند. انحراف‌های انتهایی نیز در تفسیر بازه‌های پیش‌بینی مبتنی بر تقریب نرمال باید مورد توجه قرار گیرند.

۳-۸- ارزیابی عملکرد پیش‌بینی دو مدل مرجع

پس از بررسی کفایت تشخیصی، عملکرد پیش‌بینی دو مدل مرجع $SARIMA(1,0,0)(1,1,2)_{12}$ و $SARIMA(1,1,2)(1,1,2)_{12}$ در دو دوره ارزیابی پایانی مقایسه شد. در ارزیابی ۱۲ماهه، هر مدل با داده‌های ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۳ برازش شد و دبی ژانویه تا دسامبر ۲۰۲۴ را پیش‌بینی کرد. در ارزیابی ۲۴ماهه، دوره آموزش ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۲ بود و پیش‌بینی‌ها ژانویه ۲۰۲۳ تا دسامبر ۲۰۲۴ را پوشش دادند. در هر سناریو، پیش‌بینی‌ها از مبدأ ثابتِ انتهای دوره آموزش و بدون به‌روزرسانی با مشاهدات دوره هدف تولید شدند. مطابق بخش ۲-۸، انتخاب ساختار و تنظیمات مدل‌ها از ارزیابی این دوره‌ها جدا بود و مشاهدات هدف در انتخاب مدل یا برآورد پارامترهای پیش‌بینی متناظر استفاده نشدند.

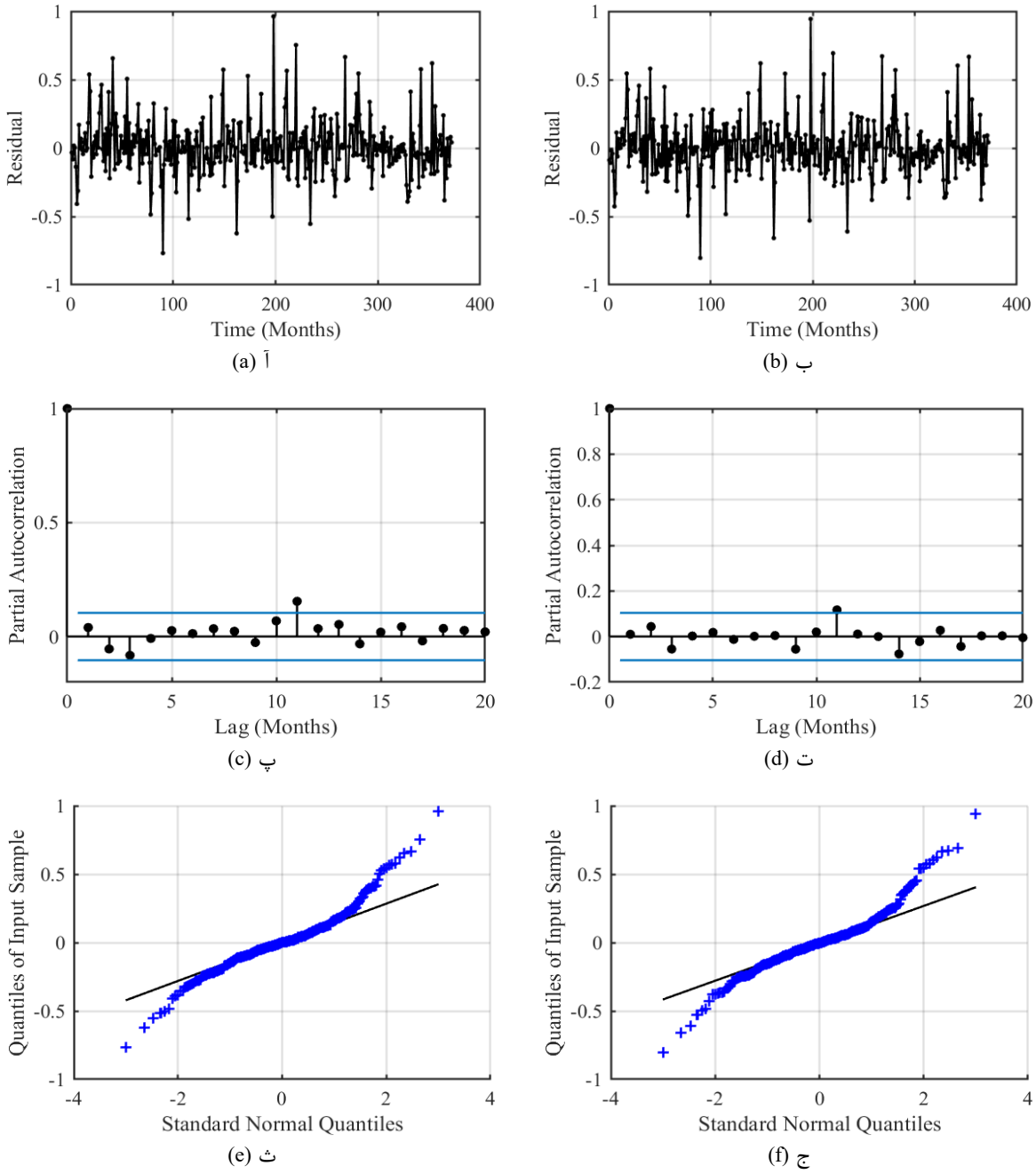
در ارزیابی ۱۲ماهه، مسیر پیش‌بینی دو مدل به یکدیگر نزدیک بود و هر دو، زمان وقوع بیشینه دبی ماهانه در ژوئن و افت جریان پس از آن را بازنمایی کردند (شکل ۵). بااین‌حال، دبی آوریل کمتر از مقدار مشاهده‌شده و دبی ماه‌های مه تا ژوئیه بیشتر از مشاهدات پیش‌بینی شد. بیش‌برآوردی قله ژوئن برای مدل دارای فقط تفاضل فصلی و مدل دارای تفاضل ترکیبی به‌ترتیب ۱۸/۲۶ و ۱۵/۴۳ درصد بود. بنابراین، بازنمایی زمان وقوع قله با برآورد دقیق دامنه آن همراه نبود.

در ارزیابی ۲۴ ماهه، تفاوت پیش‌بینی دو مدل در ماه‌های پُردبی آشکارتر بود (شکل ۶). هر دو مدل دبی مه و ژوئن ۲۰۲۳ را کم‌برآورد کردند، اما پیش‌بینی مدل دارای فقط تفاضل فصلی به مشاهدات نزدیک‌تر بود. کم‌برآوردی قله ژوئن ۲۰۲۳ برای مدل‌های دارای فقط تفاضل فصلی و تفاضل ترکیبی به‌ترتیب ۱۷/۴۴ و ۲۷/۶۱ درصد بود. در مقابل، برای قله ژوئن ۲۰۲۴ در همین سناریو، مدل نخست ۱۱/۸۹ درصد بیش‌برآوردی و مدل دوم ۶/۲۴ درصد کم‌برآوردی نشان داد. ازاین‌رو، جهت و اندازه خطای قله‌ها به ساختار مدل، ماه هدف و مبدأ پیش‌بینی وابسته بود و کم‌برآوردی را نمی‌توان به همه قله‌ها تعمیم داد.

شاخص‌های عملکرد دو مدل و دو روش مبنا در جدول ۱۰ ارائه شده‌اند. روش‌های مبنا شامل ساده فصلی و میانگین ماهانه بلندمدت بودند و با همان دوره‌های آموزش و مشاهدات هدف ارزیابی شدند. در روش ساده فصلی، آخرین چرخه ۱۲ماهه دوره آموزش برای طول دوره هدف تکرار شد؛ در روش میانگین ماهانه بلندمدت نیز میانگین حسابی دبی هر ماه تقویمی صرفا از داده‌های آموزش و در مقیاس اصلی دبی محاسبه شد. هیچ‌یک از روش‌های مبنا در طول ارزیابی با مشاهدات جدید به‌روزرسانی نشدند.

در جدول ۱۰، RMSE و MAE برحسب مترمکعب بر ثانیه، SMAPE برحسب درصد و NSE بدون بعد گزارش شده‌اند.

در دوره ۱۲ماهه، مدل دارای تفاضل ترکیبی در هر چهار شاخص عملکرد بهتری از مدل دارای فقط تفاضل فصلی داشت؛ در دوره ۲۴ماهه، ترتیب عملکرد دو مدل معکوس شد. هر دو ساختار در هر دو دوره، نسبت به روش ساده



(a) آ (b) ب
(c) پ (d) ت
(e) ث (f) ج

Fig. 4 Comparison of residual diagnostics in the logarithmic scale for models fitted to the full January 1994–December 2024 series. The left column corresponds to SARIMA(1,0,0)(1,1,2)$_{12}$ and the right column to SARIMA(1,1,2)(1,1,2)$_{12}$: (a, b) residual time series; (c, d) partial autocorrelation functions; and (e, f) normal Q–Q plots. Horizontal lines on either side of zero in panels (c) and (d) indicate approximate 95% reference bounds.

شکل ۴ مقایسه نمودارهای تشخیصی باقیمانده‌ها در مقیاس لگاریتمی، حاصل از برازش کل دوره ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴. ستون چپ مربوط به SARIMA(1,0,0)(1,1,2)$_{12}$ و ستون راست مربوط به SARIMA(1,1,2)(1,1,2)$_{12}$ است: (آ و ب) تغییرات زمانی باقیمانده‌ها؛ (پ و ت) توابع خودهمبستگی جزئی؛ (ث و ج) نمودارهای چندک-چندک نرمال. خطوط افقی دو سوی صفر در پنل‌های پ و ت، حدود مرجع تقریبی ۹۵ درصد را نشان می‌دهند.

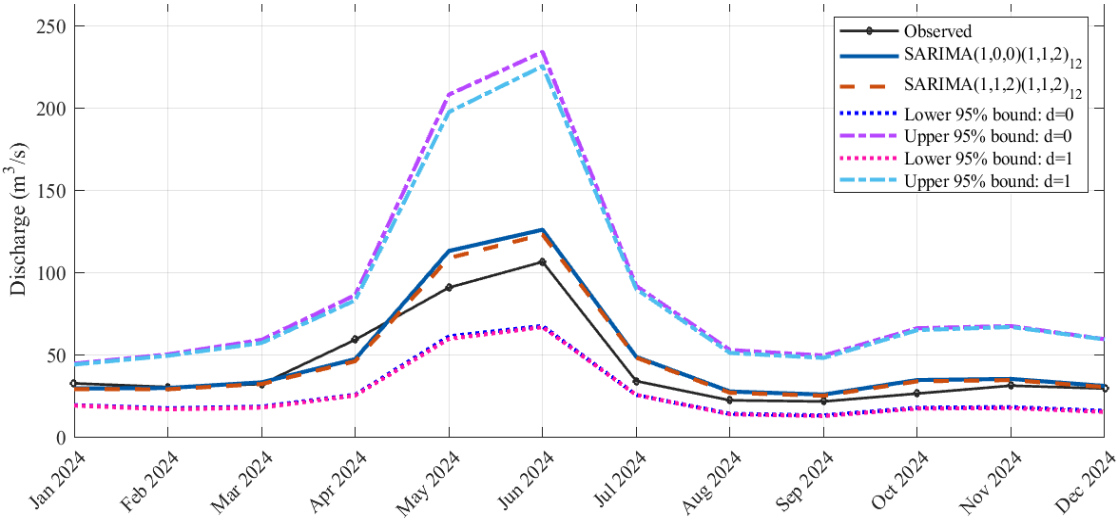


Fig. 5 Observed discharge, point forecasts, and pointwise 95% prediction interval bounds for the two reference SARIMA models in 2024. Models were fitted to January 1994–December 2023 data, with forecasts issued from the fixed December 2023 origin.

شکل ۵ دبی مشاهده‌شده، پیش‌بینی‌های نقطه‌ای و حدود بازه‌های پیش‌بینی نقطه‌ای ۹۵ درصد دو مدل مرجع SARIMA برای سال ۲۰۲۴، حاصل از برازش به داده‌های ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۳ و پیش‌بینی از مبدأ ثابت دسامبر ۲۰۲۳.

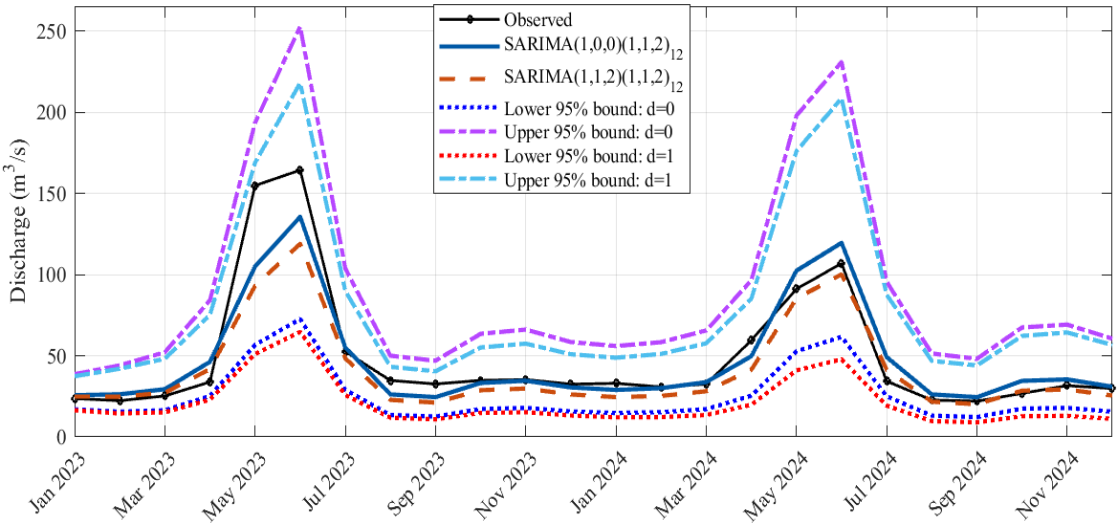


Fig. 6 Observed discharge, point forecasts, and pointwise 95% prediction interval bounds for the two reference SARIMA models in 2023–2024. Models were fitted to January 1994–December 2022 data, with forecasts issued from the fixed December 2022 origin.

شکل ۶ دبی مشاهده‌شده، پیش‌بینی‌های نقطه‌ای و حدود بازه‌های پیش‌بینی نقطه‌ای ۹۵ درصد دو مدل مرجع ساریما برای سال‌های ۲۰۲۴-۲۰۲۳، حاصل از برازش به داده‌های ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۲ و پیش‌بینی از مبدأ ثابت دسامبر ۲۰۲۲.

فصلی، RMSE،MAE و SMAPE کمتر و NSE بیشتری داشتند. برتری نسبت به میانگین ماهانه بلندمدت به ساختار و شاخص ارزیابی وابسته بود: هر دو مدل در دوره ۱۲ماهه و مدل دارای فقط تفاضل فصلی در دوره ۲۴ماهه در هر چهار شاخص بهتر عمل کردند، اما مدل دارای تفاضل ترکیبی در دوره ۲۴ماهه، با وجود MAE و SMAPE کمتر، RMSE بیشتر و NSE پایین‌تری داشت. این نتایج، همراه با تحلیل بخش ۳-۷، نشان می‌دهند که وضعیت مطلوب‌تر آزمون‌های باقیمانده لزوما با خطای پیش‌بینی کمتر همراه نیست.

مقایسه خطاهای دو دوره باید با توجه به تفاوت مشاهدات هدف، مبدأ پیش‌بینی و داده‌های آموزش انجام شود. مطابق جدول ۱۱، سال ۲۰۲۳ نسبت به سال ۲۰۲۴ میانگین، انحراف معیار و بیشینه دبی بیشتری داشت. در جدول ۱۱، N تعداد مشاهدات ماهانه و SD انحراف معیار نمونه‌ای است. تمام آماره‌های دبی برحسب مترمکعب بر ثانیه گزارش شده‌اند.

تفکیک سالانه پیش‌بینی ۲۴ماهه نشان داد که در هر دو مدل، بخش غالب مجموع مربع خطاها در سال نخست دوره هدف متمرکز بود. سهم سال ۲۰۲۳ برای مدل دارای فقط تفاضل فصلی ۸۳/۶۳ درصد و برای مدل دارای تفاضل ترکیبی ۹۱/۲۲ درصد بود. مقایسه شاخص‌های سالانه نیز نشان داد که مدل نخست در سال ۲۰۲۳ در هر چهار شاخص بهتر عمل کرد، اما در سال ۲۰۲۴، مدل دارای تفاضل ترکیبی RMSE و MAE کمتر و NSE بیشتری داشت؛ در مقابل، SMAPE مدل دارای فقط تفاضل فصلی کمتر بود (جدول ۱۲).

در جدول ۱۲، دوره آموزش همه برازش‌ها از ژانویه ۱۹۹۴ آغاز می‌شود. گام پیش‌بینی برحسب ماه است. نتایج سال‌های ۲۰۲۳ و ۲۰۲۴ از مبدأ دسامبر ۲۰۲۲، تفکیک همان پیش‌بینی ۲۴ماهه‌اند و بدون برازش مجدد یا به‌روزرسانی بین دو سال به دست آمده‌اند.بزرگ‌ترین خطای ماهانه هر دو مدل در پیش‌بینی ۲۴ماهه به مه ۲۰۲۳، گام پنجم پیش‌بینی، مربوط بود. در برابر دبی مشاهده‌شده ۱۵۴/۸۰۲ مترمکعب بر ثانیه، پیش‌بینی مدل دارای فقط تفاضل فصلی ۱۰۴/۹۷۸ و پیش‌بینی مدل دارای

جدول ۱۰ مقایسه عملکرد دو مدل مرجع SARIMA با روش ساده فصلی و میانگین ماهانه بلندمدت در دوره‌های ارزیابی ۱۲ و ۲۴ماهه.

Table 10 Performance of the two reference SARIMA models, seasonal naive forecasts, and monthly climatological means over the 12- and 24-month evaluation periods.

Evaluation period	Model or method	RMSE (m³/s)	MAE (m³/s)	SMAPE (%)	NSE
12 months: 2024	Seasonal naive	27.220	18.913	33.90	−0.044
12 months: 2024	Monthly climatological mean	17.746	11.602	20.03	0.556
12 months: 2024	SARIMA(1,0,0)(1,1,2)$_{12}$	10.709	8.045	16.10	0.838
12 months: 2024	SARIMA(1,1,2)(1,1,2)$_{12}$	9.501	7.209	14.80	0.873
24 months: 2023–2024	Seasonal naive	23.207	13.040	22.32	0.648
24 months: 2023–2024	Monthly climatological mean	16.046	10.736	19.43	0.831
24 months: 2023–2024	SARIMA(1,0,0)(1,1,2)$_{12}$	13.480	8.230	15.09	0.881
24 months: 2023–2024	SARIMA(1,1,2)(1,1,2)$_{12}$	17.026	9.734	18.74	0.810

جدول ۱۱ آمار توصیفی دبی مشاهده‌شده در دوره‌های هدف و سال ۲۰۲۳.

Table 11 Descriptive statistics of observed discharge for the target periods and the year 2023.

Period	N	Mean (m³/s)	Sample SD (m³/s)	Minimum (m³/s)	Maximum (m³/s)
January–December 2024	12	43.458	27.824	22.046	106.858
January 2023–December 2024	24	48.686	39.928	22.046	164.342
January–December 2023	12	53.915	49.996	22.348	164.342

جدول ۱۲ مقایسه عملکرد دو مدل مرجع در سال ۲۰۲۴ از دو مبدأ پیش‌بینی و در تفکیک سالانه پیش‌بینی ۲۴ماهه.

Table 12 Performance of the two reference models for 2024 from two forecast origins and for annual subsets of the 24-month forecast.

SARIMA structure	Target year	Forecast origin	Forecast steps	RMSE (m³/s)	MAE (m³/s)	SMAPE (%)	NSE
(1,0,0)(1,1,2)$_{12}$	2024	Dec 2023	1–12	10.709	8.045	16.10	0.838
(1,1,2)(1,1,2)$_{12}$	2024	Dec 2023	1–12	9.501	7.209	14.80	0.873
(1,0,0)(1,1,2)$_{12}$	2023	Dec 2022	1–12	17.434	10.373	16.85	0.867
(1,1,2)(1,1,2)$_{12}$	2023	Dec 2022	1–12	22.996	13.850	22.97	0.769
(1,0,0)(1,1,2)$_{12}$	2024	Dec 2022	13–24	7.713	6.087	13.34	0.916
(1,1,2)(1,1,2)$_{12}$	2024	Dec 2022	13–24	7.136	5.619	14.51	0.928

تفاضل ترکیبی ۹۳/۰۴۷ مترمکعب بر ثانیه بود. این ماه به‌ترتیب ۵۶/۹۲ و ۵۴/۸۲ درصد مجموع مربع خطاهای دو مدل را تشکیل داد. بنابراین، سهم بالای خطاهای سال ۲۰۲۳ با کم‌برآوردی جریان‌های بزرگ در ابتدای دوره هدف همراه بود.

برای مشاهدات یکسان سال ۲۰۲۴، هر دو مدل از مبدأ دسامبر ۲۰۲۲، نسبت به مبدأ دسامبر ۲۰۲۳، RMSE و MAE کمتر و NSE بیشتری داشتند. این نتیجه نشان می‌دهد که در ارزیابی حاضر، فاصله بیشتر از مبدأ پیش‌بینی الزاما با خطای بیشتر همراه نبوده است. بااین‌حال، به دلیل تفاوت داده‌های آموزش و پارامترهای برآوردشده در دو مبدأ، این مقایسه اثر مستقل افق پیش‌بینی را از سایر عوامل جدا نمی‌کند.

برای بررسی عملکرد در شرایط مختلف جریان، ماه‌های کم‌آب و پُرآب بر اساس صدک‌های ۲۵ و ۷۵ دبی دوره آموزش هر سناریو دسته‌بندی شدند. آستانه کم‌آبی در هر دو سناریو ۲۸/۶۱۲ مترمکعب بر ثانیه و آستانه پُرآبی در ارزیابی‌های ۱۲ و ۲۴ماهه به‌ترتیب ۵۱/۸۴۷ و ۵۱/۹۹۷ مترمکعب بر ثانیه بود. به دلیل یکسان‌بودن داده‌های آموزش و مشاهدات هدف، دسته‌بندی ماه‌ها برای دو مدل یکسان بود. شاخص‌های خطا و سوگیری حجمی در جدول ۱۳ ارائه شده‌اند.

در جدول ۱۳، Nتعداد ماه‌ها،MRE میانگین خطای نسبی علامت‌دار، MAPE میانگین قدرمطلق خطای نسبی و B_v سوگیری حجمی است. مقادیر مثبت MRE و B_v نشان‌دهنده کم‌برآوردی و مقادیر منفی بیانگر بیش‌برآوردی‌اند. گروه «همه ماه‌ها» شامل جریان‌های میانی نیز می‌شود. حجم‌ها با لحاظ تعداد واقعی روزهای هر ماه محاسبه شده‌اند. در ارزیابی ۱۲ماهه، مدل دارای تفاضل ترکیبی در هر دو گروه کم‌آب و پُرآب، RMSE، MAE و MAPE کمتر و قدرمطلق سوگیری حجمی کوچک‌تری داشت. هر دو مدل تمام ماه‌های کم‌آب را بیش‌برآورد کردند، اما خطاهای گروه پُرآب هر دو علامت را داشتند؛ بنابراین، MRE منفی این گروه بیانگر بیش‌برآوردی همه ماه‌های آن نیست.

در ارزیابی ۲۴ماهه، عملکرد دو مدل میان گروه‌های جریان متفاوت بود. مدل دارای فقط تفاضل فصلی در گروه پُرآب خطاهای مطلق و نسبی کمتر و سوگیری حجمی کوچک‌تری داشت، درحالی‌که مدل دارای تفاضل ترکیبی در گروه کم‌آب بهتر عمل کرد. در مدل نخست، چهار ماه از شش ماه پُرآب کم‌برآورد و دو ماه بیش‌برآورد شدند؛ در مدل دوم، هر شش ماه پُرآب کم‌برآورد شدند. این تفاوت توضیح می‌دهد که چرا MRE مدل نخست در گروه پُرآب از MAPE آن کمتر بود، اما این دو شاخص برای مدل دوم برابر بودند.

جدول ۱۳ شاخص‌های خطا و سوگیری حجمی دو مدل مرجع برای کل دوره و ماه‌های کم‌آب و پُرآب در دو سناریوی ارزیابی.

Table 13 Error metrics and volume bias of the two reference models for the full evaluation periods and their low- and high-flow months.

Evaluation period	SARIMA structure	Flow group	N	RMSE (m³/s)	MAE (m³/s)	MRE (%)	MAPE (%)	B_v (%)
12 months: 2024	$(1,0,0)(1,1,2)_{12}$	All	12	10.709	8.045	−12.44	17.63	−12.71
12 months: 2024	$(1,1,2)(1,1,2)_{12}$	All	12	9.501	7.209	−9.79	15.93	−9.86
12 months: 2024	$(1,0,0)(1,1,2)_{12}$	Low	3	6.157	5.921	−24.39	24.39	−24.85
12 months: 2024	$(1,1,2)(1,1,2)_{12}$	Low	3	5.428	5.164	−21.20	21.20	−21.68
12 months: 2024	$(1,0,0)(1,1,2)_{12}$	High	3	18.386	17.842	−7.63	20.81	−11.79
12 months: 2024	$(1,1,2)(1,1,2)_{12}$	High	3	15.968	15.835	−4.40	19.02	−8.44
24 months: 2023–2024	$(1,0,0)(1,1,2)_{12}$	All	24	13.480	8.230	−3.46	15.36	2.64
24 months: 2023–2024	$(1,1,2)(1,1,2)_{12}$	All	24	17.026	9.734	10.78	16.94	16.22
24 months: 2023–2024	$(1,0,0)(1,1,2)_{12}$	Low	6	4.370	3.965	−16.32	16.32	−16.66
24 months: 2023–2024	$(1,1,2)(1,1,2)_{12}$	Low	6	1.758	1.691	−2.88	7.15	−3.05
24 months: 2023–2024	$(1,0,0)(1,1,2)_{12}$	High	6	24.809	19.111	6.19	15.83	9.86
24 months: 2023–2024	$(1,1,2)(1,1,2)_{12}$	High	6	32.400	23.684	19.80	19.80	22.60

سوگیری حجمی کل دوره در هر دو مدل از بیش‌برآوردی در ارزیابی ۱۲ماهه به کم‌برآوردی در ارزیابی ۲۴ماهه تغییر کرد. در مدل دارای فقط تفاضل فصلی، MRE کل دوره ۲۴ماهه منفی بود، درحالی‌که سوگیری حجمی مثبت به دست آمد. این تفاوت از وزن‌دهی متفاوت دو شاخص ناشی می‌شود. MRE ، میانگین خطاهای نسبی ماهانه است، اما سوگیری حجمی به اندازه دبی و طول هر ماه وابسته است. ازاین‌رو، جهت میانگین خطای نسبی لزوما با جهت خطای حجم کل دوره یکسان نیست. همچنین، به دلیل تعداد محدود مشاهدات هر گروه و ماهیت نسبی دسته‌بندی جریان، این نتایج به‌تنهایی برای تایید کاربرد عملیاتی مدل‌ها در مدیریت کم‌آبی یا پیش‌بینی جریان‌های بزرگ کافی نیستند.

حدود بازه‌های پیش‌بینی نقطه‌ای ۹۵ درصد هر دو مدل در شکل های ۵ و ۶ و ارزیابی عددی آن‌ها در جدول ۱۴ ارائه شده‌اند. تمام مشاهدات هر دو دوره داخل بازه‌های متناظر هر دو مدل قرار گرفتند. با پوشش تجربی یکسان، میانگین پهنای بازه‌های مدل دارای تفاضل ترکیبی در هر دو دوره کمتر بود.در جدول ۱۴، بازه‌ها نقطه‌ای‌اند و پوشش هم‌زمان کل مسیر پیش‌بینی را بیان نمی‌کنند. حدود با تبدیل $\exp(y)-1$ از مقیاس $\ln(Q+1)$ به مقیاس دبی بازگردانده شده‌اند. در هیچ‌یک از ارزیابی‌ها، حد پایین به محدودشدن در صفر نیاز نداشت.

بازه‌ها پس از تبدیل معکوس در مقیاس اصلی دبی نامتقارن بودند و در ماه‌های دارای پیش‌بینی دبی بیشتر، پهنای بیشتری داشتند. مشاهدات پُرآب مه و ژوئن ۲۰۲۳ نیز، با وجود کم‌برآوردی پیش‌بینی نقطه‌ای هر دو مدل، داخل بازه‌های متناظر قرار گرفتند. بااین‌حال، پوشش تجربی ۱۰۰ درصد در این دو دوره کوتاه، به‌تنهایی انطباق پوشش بلندمدت با سطح اسمی ۹۵ درصد را تایید نمی‌کند. تفسیر این نتیجه باید هم‌زمان با پهنای بازه‌ها و انحراف توزیع باقیمانده‌ها از نرمال، گزارش‌شده در بخش ۳-۷، انجام شود.

۳-۹- پیش‌بینی دبی ماهانه برای دوره ژانویه ۲۰۲۵ تا دسامبر ۲۰۲۶

برای بررسی پیش‌بینی‌های فراتر از انتهای سری مورد مطالعه، دو مدل مرجع $SARIMA(1,0,0)(1,1,2)_{12}$ و $SARIMA(1,1,2)(1,1,2)_{12}$ با استفاده از کل داده‌های ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴ دوباره برازش شدند. سپس پیش‌بینی‌های ۲۴ماهه از مبدأ ثابت دسامبر ۲۰۲۴، بدون به‌روزرسانی با مشاهدات دوره هدف، تولید شدند. پیش‌بینی‌های نقطه‌ای و حدود بازه‌های پیش‌بینی نقطه‌ای ۹۵ درصد هر دو مدل در شکل ۷ ارائه شده‌اند.

هر دو مدل افزایش دبی در بهار، وقوع بیشینه در ژوئن و کاهش جریان پس از آن را پیش‌بینی کردند. زمان‌بندی تغییرات فصلی دو مسیر مشابه بود، اما دامنه پیش‌بینی‌ها تفاوت داشت؛ به‌ویژه در ماه‌های مه و ژوئن هر دو سال، مدل دارای فقط تفاضل فصلی مقادیر بیشتری نسبت به مدل دارای تفاضل ترکیبی پیش‌بینی کرد. این اختلاف بیانگر حساسیت دامنه پیش‌بینی به ساختار مدل است و بدون مقایسه با مشاهدات دوره هدف، نمی‌توان آن را نشانه دقت بیشتر یکی از دو مدل دانست.

جدول ۱۴ پوشش تجربی و میانگین پهنای بازه‌های پیش‌بینی نقطه‌ای ۹۵ درصد دو مدل مرجع در دوره‌های ارزیابی پایانی.

Table 14 Empirical coverage and mean width of pointwise 95% prediction intervals for the two reference models over the final evaluation periods.

Evaluation period	SARIMA structure	Nominal coverage (%)	Covered observations	Empirical coverage (%)	Mean width (m^3/s)	Below lower bound	Above upper bound
12 months: 2024	$(1,0,0)(1,1,2)_{12}$	95	12/12	100	62.893	0	0
12 months: 2024	$(1,1,2)(1,1,2)_{12}$	95	12/12	100	60.726	0	0
24 months: 2023–2024	$(1,0,0)(1,1,2)_{12}$	95	24/24	100	64.749	0	0
24 months: 2023–2024	$(1,1,2)(1,1,2)_{12}$	95	24/24	100	59.079	0	0

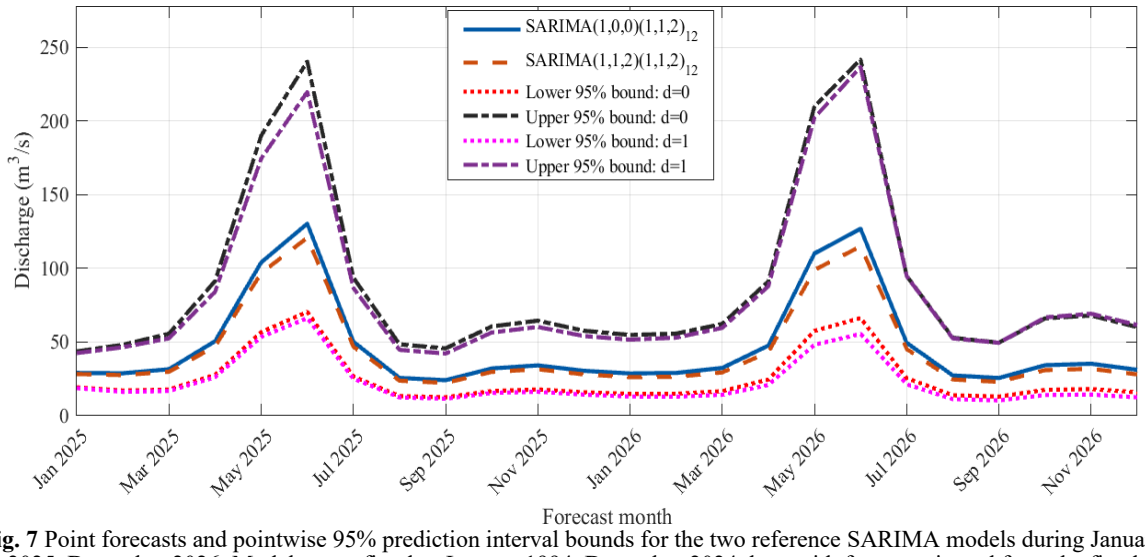


Fig. 7 Point forecasts and pointwise 95% prediction interval bounds for the two reference SARIMA models during January 2025–December 2026. Models were fitted to January 1994–December 2024 data, with forecasts issued from the fixed December 2024 origin.

شکل ۷ پیش‌بینی‌های نقطه‌ای و حدود بازه‌های پیش‌بینی نقطه‌ای ۹۵ درصد دو مدل مرجع ساریما برای ژانویه ۲۰۲۵ تا دسامبر ۲۰۲۶، حاصل از برازش به داده‌های ژانویه ۱۹۹۴ تا دسامبر ۲۰۲۴ و پیش‌بینی از مبدأ ثابت دسامبر ۲۰۲۴.

بازه‌های پیش‌بینی دو مدل هم‌پوشانی گسترده‌ای داشتند و پس از تبدیل معکوس از مقیاس $\ln(Q+1)$ به مقیاس دبی، نسبت به پیش‌بینی نقطه‌ای نامتقارن بودند. پهنای مطلق بازه‌ها در ماه‌های دارای پیش‌بینی دبی بیشتر افزایش یافت. این ویژگی به واریانس خطای پیش‌بینی در مقیاس لگاریتمی و تبدیل غیرخطی به مقیاس اصلی وابسته است؛ بنابراین، پهنای بیشتر بازه در ماه‌های پُردبی را نمی‌توان به‌تنهایی نشانه افزایش عدم‌قطعیت نسبی دانست. همچنین، بازه‌های ارائه‌شده برای هر ماه به‌صورت نقطه‌ای محاسبه شده‌اند و پوشش هم‌زمان کل مسیر ۲۴ماهه را بیان نمی‌کنند.

پیش‌بینی‌های این دوره، برآوردهای مشروط به ساختار هر مدل و اطلاعات موجود تا دسامبر ۲۰۲۴ هستند. در تحلیل حاضر، مشاهدات دوره هدف برای ارزیابی صحت پیش‌بینی‌ها یا پوشش تجربی بازه‌ها استفاده نشده‌اند. ازاین‌رو، مشابهت چرخه فصلی دو مسیر یا هم‌پوشانی بازه‌های آن‌ها، صحت مقادیر پیش‌بینی‌شده را تایید نمی‌کند. با توجه به وابستگی عملکرد مدل‌ها به دوره و شرایط جریان در بخش ۳-۸، این نتایج نیز مبنای معرفی یک مدل به‌عنوان ساختار برتر در دوره ۲۰۲۶-۲۰۲۵ نیستند. قابلیت استفاده عملیاتی از این پیش‌بینی‌ها در برنامه‌ریزی منابع آب در این پژوهش ارزیابی نشده است.

۳-۱۰- بحث

نتایج نشان داد که ترجیح یک ساختار بر اساس معیارهای اطلاعاتی الزاما به رفع وابستگی زمانی باقیمانده‌ها منجر نمی‌شود. در گروه آریما با $d=0$، معیارهای AIC و BIC به‌ترتیب ساختارهای (۲,۰,۲) و (۱,۰,۱) را ترجیح دادند؛ در گروه $d=1$، هر دو معیار ساختار (۲,۱,۲) را پیشنهاد کردند. بااین‌حال، فرض صفر نبود خودهمبستگی باقیمانده‌ها برای هر سه ساختار تا وقفه‌های ۱۲ و ۲۴ رد شد (جدول ۸). در میان دو مدل مرجع ساریما، این فرض برای مدل دارای تفاضل ترکیبی در هیچ‌یک از دو آزمون رد نشد، اما برای مدل دارای فقط تفاضل فصلی تا وقفه ۱۲ رد شد (جدول ۹). این تفاوت، وضعیت تشخیصی مطلوب‌تر مدل دارای تفاضل ترکیبی را در آزمون‌های مذکور نشان می‌دهد؛ بااین‌حال، مقایسه حاضر بر باقیمانده‌های برازش کل دوره متکی است و به‌تنهایی برتری پیش‌بینی ساریما نسبت به آریما یا ضرورت تفاضل‌گیری فصلی را اثبات نمی‌کند. همچنین، عدم رد فرض صفر به معنای اثبات استقلال باقیمانده‌ها نیست و رد آزمون تا وقفه ۱۲ نیز محل یا ماهیت فصلی وابستگی باقی‌مانده را مشخص نمی‌کند.

کاربرد ساختارهای فصلی برای پیش‌بینی دبی ماهانه در مطالعات پیشین نیز بررسی شده است. Alonso Brito و همکاران (۲۰۲۱) عملکرد مدل‌های ساریما و هولت-وینترز را مقایسه کردند. Sheikholeslami و همکاران (۲۰۲۴) نیز با استفاده از داده‌های ماهانه سال‌های ۲۰۲۳-۲۰۰۶ ایستگاه پل زغال در رودخانه چالوس، مدل $SARIMA(2,0,0)(0,1,1)_{12}$ را انتخاب کردند و گزارش دادند که بازنمایی تغییرات فصلی با محدودیت در پیش‌بینی جریان‌های بسیار بزرگ همراه بوده است. این یافته با تمایز مشاهده‌شده در پژوهش حاضر میان بازنمایی زمان‌بندی فصلی و دقت برآورد دامنه قله‌ها همسو است. بااین‌حال، تفاوت ایستگاه، پیش‌پردازش و طرح ارزیابی مانع تعمیم مستقیم عملکرد آن مدل به رودخانه سالمون می‌شود. همچنین، استفاده از ساختاری با $d=0$ و $D=1$ در مطالعه چالوس، به‌تنهایی ضرورت یا کفایت همین مراتب تفاضل‌گیری را برای داده‌های پژوهش حاضر اثبات نمی‌کند.

تحلیل پیچیدگی، تفاوت میان کفایت تشخیصی و دقت پیش‌بینی را روشن‌تر کرد. در گروه دارای فقط تفاضل فصلی، مدل مرجع کمترین امتیاز تجمیعی خطای اعتبارسنجی را در مجموعه بررسی‌شده داشت، هرچند فرض صفر آزمون لیونگ-باکس تا وقفه ۱۲ در هر سه مبدأ رد شد. در گروه دارای تفاضل ترکیبی، مدل مرجع از نظر معیارهای اطلاعاتی و آزمون‌های باقیمانده مطلوب‌تر بود، اما چندین ساختار ساده‌تر خطای اعتبارسنجی کمتری داشتند. بنابراین، پیچیدگی بیشتر این مدل از منظر کاهش خطای پیش‌بینی در مجموعه بررسی‌شده توجیه نشد. این نتیجه نشان می‌دهد که انتخاب ساختار تنها بر اساس معیارهای اطلاعاتی یا عدم رد آزمون خودهمبستگی، برای تعیین مدل کم‌خطاتر کافی نیست. تعداد محدود مبدأهای اعتبارسنجی و هم‌پوشانی افق‌ها نیز مبنای ادعای برتری آماری قطعی یک ساختار را فراهم نمی‌کنند.

شواهد موجود ضرورت هم‌زمان تفاضل‌گیری غیرفصلی و فصلی را اثبات نمی‌کنند. پس از تفاضل فصلی به‌تنهایی نیز فرض ریشه واحد در تمام مشخصات آزمون ADF بررسی‌شده رد شد و مدل مرجع این گروه، میانگین خطای اعتبارسنجی کمتری از مدل مرجع دارای تفاضل ترکیبی داشت. بااین‌حال، دو مدل علاوه بر مرتبه تفاضل‌گیری غیرفصلی، در مرتبه میانگین متحرک غیرفصلی نیز متفاوتاند؛ ازاین‌رو، تفاوت عملکرد یا رفتار باقیمانده‌های آن‌ها را نمی‌توان صرفا به افزودن تفاضل غیرفصلی نسبت داد. مقایسه آریما با ساریما نیز اثر تفاضل فصلی را از اثر جمله‌های خودرگرسیونی و میانگین متحرک فصلی جدا نمی‌کند. بنابراین، $D=1$ در این مطالعه مشخصه ساختارهای فصلی بررسی‌شده است و ضرورت آن با آزمون اختصاصی ریشه واحد فصلی اثبات نشده است.

ارزیابی‌های پایانی نیز از برتری یکنواخت یکی از دو مدل پشتیبانی نکردند. مدل دارای تفاضل ترکیبی در دوره ۱۲ماهه در هر چهار شاخص بهتر عمل کرد، اما در دوره ۲۴ماهه مدل دارای فقط تفاضل فصلی برتر بود. هر دو ساختار از روش ساده فصلی بهتر عمل کردند؛ بنابراین، عملکرد آن‌ها در این دوره‌ها صرفا معادل تکرار آخرین چرخه سالانه نبود. در مقابل، برتری نسبت به میانگین ماهانه بلندمدت به ساختار و شاخص وابسته بود. مدل دارای تفاضل ترکیبی در دوره ۲۴ماهه، با وجود MAE و SMAPE کمتر،RMSE بیشتر و NSE پایین‌تری از این مبنا داشت. این اختلاف با حساسیت بیشتر معیارهای مبتنی بر مربع خطا به خطاهای بزرگ سازگار است و نشان می‌دهد که نتیجه مقایسه باید با ذکر شاخص و دوره ارزیابی بیان شود.

کم‌برآوردی جریان‌های بزرگ سال ۲۰۲۳، در کنار بیش‌برآوردی برخی قله‌های سال ۲۰۲۴، نشان داد که بازنمایی زمان‌بندی فصلی با دقت یکسان در دامنه جریان همراه نیست. محدودیت مدل‌های سری زمانی در پیش‌بینی برخی جریان‌های بزرگ در مطالعات پیشین نیز گزارش شده است. Kenyi و Yamamoto (2024) برای مدل‌های منفرد ساریما و پروفت، انحراف پیش‌بینی‌ها در دوره‌های پُرجریان را گزارش کردند؛ Kang و همکاران (۲۰۲۵) نیز در بررسی رویکرد اس‌تی‌ال-ساریما، افزایش پراکندگی مولفه باقیمانده در فصل سیلاب و محدودیت پیش‌بینی جریان‌های بزرگ را مطرح کردند. این شواهد با ضرورت تفکیک بازنمایی الگوی فصلی از دقت پیش‌بینی قله‌ها همسو هستند، اما تفاوت حوضه‌ها و طرح‌های ارزیابی اجازه نمی‌دهد جهت یا اندازه خطاهای آن مطالعات مستقیما به

رودخانه سالمون تعمیم داده شود.

در تفسیر دامنه پیش‌بینی‌ها، روش بازگشت به مقیاس اصلی نیز باید در نظر گرفته شود. در هر دو مدل، تبدیل معکوس مستقیم بدون تصحیح سوگیری انجام شد. تحت فرض توزیع نرمال پیش‌بینی در مقیاس لگاریتمی و پیش از اعمال قید نامنفی‌بودن، این تبدیل میانه شرطی دبی را به دست می‌دهد که در صورت مثبت‌بودن واریانس پیش‌بینی، از میانگین شرطی کمتر است. بنابراین، روش تبدیل معکوس می‌تواند در پایین‌تر بودن پیش‌بینی‌ها نسبت به برآورد میانگین شرطی سهم داشته باشد؛ بااین‌حال، کم‌برآوردی قله‌های مشاهده‌شده را نمی‌توان صرفا به این عامل یا ساختار ساریما نسبت داد. چون پیش‌بینی‌های دارای تصحیح سوگیری در این پژوهش مقایسه نشده‌اند، سهم این عامل در خطاها مشخص نیست و بهبود دقت با اعمال تصحیح نیز تایید نشده است.

مدل‌های بررسی‌شده تک‌متغیره‌اند و بارش، دما و شرایط مرتبط با برف مستقیما در آن‌ها وارد نشده‌اند. Costa و همکاران (۲۰۲۳) بهبود عملکرد ساریماکس را با افزودن بارش به‌عنوان متغیر توضیحی گزارش کردند؛ این نتیجه، بررسی اطلاعات هیدرواقلیمی را به‌عنوان یک مسیر توسعه مطرح می‌کند. بااین‌حال، اثر افزودن این متغیرها در ایستگاه سالمون آزمون نشده است و از نتایج حاضر نمی‌توان درباره میزان بهبود حاصل از آن‌ها نتیجه‌گیری کرد.

تحلیل گروه‌های جریان نشان داد که رتبه‌بندی مدل‌ها بر اساس کل دوره، همه تفاوت‌های عملکرد را آشکار نمی‌کند. در ارزیابی ۲۴ماهه، مدل دارای فقط تفاضل فصلی در گروه پُرآب بهتر عمل کرد، درحالی‌که مدل دارای تفاضل ترکیبی در گروه کم‌آب خطای کمتری داشت. در ارزیابی ۱۲ماهه نیز خطای مطلق ماه‌های کم‌آب کوچک‌تر بود، اما بیش‌برآوردی نسبی آن‌ها نشان داد که خطای مطلق کم الزاما به معنای برآورد دقیق آب در دسترس نیست. چنین رفتاری در صورت استفاده مدیریتی می‌تواند به برآورد خوش‌بینانه منابع موجود منجر شود، هرچند پیامد عملیاتی آن در این پژوهش ارزیابی نشده است. تفاوت علامت میانگین خطای نسبی و سوگیری حجمی مدل دارای فقط تفاضل فصلی در دوره ۲۴ماهه نیز ضرورت گزارش هم‌زمان خطاهای نسبی و حجمی را نشان می‌دهد. با توجه به وجود تنها سه یا شش مشاهده در هر گروه و مقیاس ماهانه داده‌ها، این نتایج برای تایید قابلیت مدل‌ها در مدیریت کم‌آبی یا پیش‌بینی اوج لحظه‌ای سیلاب کافی نیستند.

تفکیک زمانی خطاها نشان داد که افزایش RMSE کل دوره ۲۴ماهه در هر دو مدل با تمرکز خطاهای بزرگ در سال ۲۰۲۳ همراه بود. این سال، جریان‌های بزرگ‌تر و پراکندگی بیشتری داشت و کم‌برآوردی مه و ژوئن سهم عمده‌ای از مجموع مربع خطاها را تشکیل داد. در مقابل، برای مشاهدات یکسان سال ۲۰۲۴، هر دو مدل از مبدأ دورتر RMSE و MAE کمتری داشتند. بنابراین، نتایج از تفسیر افزایش خطا به‌عنوان پیامد صرفِ بلندترشدن افق پشتیبانی نمی‌کنند. مقایسه سال ۲۰۲۴، تفاوت ترکیب مشاهدات هدف را کنترل می‌کند، اما چون داده‌های آموزش و پارامترهای برآوردشده میان دو مبدأ متفاوتاند، اثر مستقل فاصله پیش‌بینی همچنان قابل تعیین نیست.

ارزیابی بازه‌ها نیز جنبه‌ای متفاوت از عملکرد مدل‌ها را نشان داد. هر دو مدل در دو دوره پایانی پوشش تجربی ۱۰۰ درصد داشتند، اما میانگین پهنای بازه‌های مدل دارای تفاضل ترکیبی کمتر بود. قرارگرفتن قله‌های کم‌برآوردشده درون بازه‌ها نشان می‌دهد که خطای پیش‌بینی نقطه‌ای و پوشش بازه باید جداگانه بررسی شوند؛ پوشش کامل به‌تنهایی نشانه دقت بالای پیش‌بینی نیست. تعداد محدود مشاهدات و انحراف انتهای توزیع باقیمانده‌ها از نرمال نیز مانع تایید پوشش بلندمدت بازه‌های مبتنی بر تقریب نرمال می‌شوند. به همین ترتیب، شباهت چرخه فصلی و هم‌پوشانی بازه‌های پیش‌بینی دوره ۲۰۲۶-۲۰۲۵ شواهدی از صحت مقادیر آن دوره محسوب نمی‌شوند، زیرا در تحلیل حاضر با مشاهدات هدف ارزیابی نشده‌اند.

ساختار و تنظیمات مدل‌ها با داده‌های پیش از ۲۰۲۳ تثبیت شدند و نتایج دوره‌های پایانی در انتخاب یا بازنگری آن‌ها دخالت نداشتند. بنابراین، عملکرد گزارش‌شده در این دوره‌ها حاصل ارزیابی ساختارهای از پیش تثبیت‌شده است. بااین‌حال، این جداسازی به‌تنهایی پایداری عملکرد در

شرایط دیگر را تضمین نمی‌کند. مطالعه به یک ایستگاه، سه مبدأ اعتبارسنجی و دو دوره ارزیابی پایانی محدود بود. اشتراک مشاهدات سال ۲۰۲۴ نیز مانع تلقی دو ارزیابی پایانی به‌عنوان تکرارهای آماری مستقل است. بررسی دوره‌ها و ایستگاه‌های دیگر برای ارزیابی پایداری عملکرد و قابلیت تعمیم یافته‌ها ضروری است.

سهم تجربی این پژوهش، مستندسازی تفاوت میان معیارهای انتخاب، تشخیص باقیمانده‌ها و عملکرد پیش‌بینی در یک طرح چندگام‌به‌جلو از مبدأ ثابت است. نتایج نشان دادند که هیچ‌یک از دو مدل مرجع در همه دوره‌ها و شرایط جریان برتر نبود و پیچیدگی بیشتر یا تشخیص مطلوب‌تر باقیمانده‌ها نیز به‌تنهایی دقت بیشتر را تضمین نکرد. ارزیابی در مبدأهای بیشتر و دوره‌های جدید، بررسی ایستگاه‌های دیگر و مقایسه با مدل‌های چندمتغیره، ترکیبی و داده‌محور می‌تواند حدود پایداری این یافته‌ها را روشن‌تر کند.

۴- نتیجه‌گیری

این پژوهش عملکرد دو مدل مرجع SARIMA را برای پیش‌بینی دبی متوسط ماهانه رودخانه سالمون از مبدأ ثابت بررسی کرد. مدل $SARIMA(1,0,0)(1,1,2)_{12}$ در اعتبارسنجی زمانی خطای کمتری از $SARIMA(1,1,2)(1,1,2)_{12}$ داشت، اما وضعیت تشخیصی باقیمانده‌های مدل دوم مطلوب‌تر بود. چندین ساختار ساده‌تر در گروه تفاضل ترکیبی نیز خطای کمتری از مدل مرجع داشتند؛ بنابراین، پیچیدگی بیشتر و تشخیص مطلوب‌تر، برتری عمومی پیش‌بینی را تایید نکردند.

در ارزیابی‌های پایانی، مدل دارای تفاضل ترکیبی در دوره ۱۲ماهه و مدل دارای فقط تفاضل فصلی در دوره ۲۴ماهه در هر چهار شاخص بهتر عمل کردند. هر دو مدل از روش ساده فصلی بهتر بودند، اما برتری نسبت به میانگین ماهانه بلندمدت یکنواخت نبود. در دوره ۲۴ ماهه، مدل دارای تفاضل ترکیبی در ماه‌های کم‌آب و مدل دارای فقط تفاضل فصلی در ماه‌های پُرآب خطای کمتری داشت. هر دو مدل جریان‌های بزرگ مه و ژوئن ۲۰۲۳ را کم‌برآورد و قله ژوئن ۲۰۲۴ را در ارزیابی ۱۲ماهه بیش‌برآورد کردند. تمرکز خطاهای بزرگ در سال ۲۰۲۳ نشان داد که افزایش خطای دوره ۲۴ماهه را نمی‌توان صرفا به طول افق نسبت داد.

تمام مشاهدات در بازه‌های پیش‌بینی نقطه‌ای ۹۵ درصد هر دو مدل قرار گرفتند، اما محدودبودن و هم‌پوشانی دوره‌ها مانع تایید پوشش بلندمدت است. داده‌های هدف در انتخاب ساختار و تنظیمات مدل‌ها دخالت نداشتند. بررسی دوره‌ها و ایستگاه‌های دیگر برای ارزیابی پایداری عملکرد و قابلیت تعمیم یافته‌ها ضروری است. نتایج حاضر برای توصیه کاربرد عملیاتی گسترده یا تایید توان پیش‌بینی قله‌های لحظه‌ای سیلاب کافی نیستند.

5- List of Symbols

ACF	Autocorrelation function
ADF	Augmented Dickey–Fuller test
AIC	Akaike information criterion
AR	Autoregressive component
ARIMA	Autoregressive integrated moving average model
B	Backshift operator
BIC	Bayesian information criterion
d	Order of non-seasonal differencing
D	Order of seasonal differencing
MA	Moving average component
MAE ($m^3 s^{-1}$)	Mean absolute error
N	Number of observations
NSE	Nash–Sutcliffe efficiency coefficient
p	Order of non-seasonal autoregression
p-value	Probability value of the statistical test
P	Order of seasonal autoregression
PACF	Partial autocorrelation function
q	Order of non-seasonal moving average
Q	Order of seasonal moving average
Q_{LB}	Ljung–Box test statistic
RMSE (m^3 $^{-1}$)	Root mean square error
S (month)	Seasonal period
SAR	Seasonal autoregressive component

	interval on the original scale
$Q_{U,t}$ ($m^3 s^{-1}$)	Upper bound of the prediction interval on the original scale
$PI_{95\%}$	95% prediction interval
$\varphi_p(B)$	Non-seasonal autoregressive polynomial of order p
$\theta q(B)$	Non-seasonal moving average polynomial of order q
$\Phi_P(B^s)$	Seasonal autoregressive polynomial of order P
$\Theta_Q(B^s)$	Seasonal moving average polynomial of order Q
MRE(%)	Mean relative error
MAPE(%)	Mean absolute percentage error
B_V(%)	Volume bias
Δt_t(s)	Duration of month t

Acknowledgments

The authors have no acknowledgments to declare.

Credit Authorship Contribution Statement

Conceptualization, H.B. and M.F.M.; Methodology, H.B. and M.F.M.; Software, H.B. and M.F.M.; Data curation, H.B. and M.F.M.; Formal analysis, H.B. and M.F.M.; Validation, H.B. and M.F.M.; Visualization, H.B. and M.F.M.; Writing – original draft, H.B. and M.F.M.; Writing – review & editing, H.B. and M.F.M.; Supervision, H.B. and M.F.M.

All authors have read and agreed to the published version of the manuscript.

Funding

This research received no external funding.

Data Availability Statement

The daily discharge data used in this study, covering January 1994 to December 2024, are publicly available from the U.S. Geological Survey's Water Data for the Nation database for station 13302500 (Salmon River at Salmon, Idaho): https://waterdata.usgs.gov/monitoring-location/USGS-13302500/. Monthly mean discharge values were calculated from these daily records following the aggregation and unit conversion procedures described in the Methodology section.

Conflicts of Interest

The authors declare no conflict of interest.

References

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control* 19(6), 716–723,

SARIMA	Seasonal autoregressive integrated moving average model
SMA	Seasonal moving average component
SMAPE (%)	Symmetric mean absolute percentage error
t (month)	time
y_t	Log-transformed discharge value at time t
$\hat{y}_t$	Fitted value or one-step-ahead forecast at time t on the logarithmic scale
ε_t	Innovation or random error of the model at time t
θ_j	Non-seasonal moving average coefficient of order j
Θ_j	Seasonal moving average coefficient of order j
φ_i	Non-seasonal autoregressive coefficient of order i
Φ_i	Seasonal autoregressive coefficient of order i
σ^2	Variance of model innovations
Q_t ($m^3 s^{-1}$)	Observed discharge at time t on the original scale
∇	First-order non-seasonal differencing operator
∇_{12}	First-order seasonal differencing operator with a 12-month period
ω_t	Series obtained after simultaneous non-seasonal and seasonal differencing
C	Model constant
$\hat{L}$	Maximum value of the likelihood function
K	Number of estimated model parameters
n	Number of observations or residuals used in the calculation
e_t	Model residual at time t
$\hat{\rho}_k$	Residual autocorrelation coefficient at lag k
k	Lag index in the Ljung–Box test
m	Maximum lag considered
df	Degrees of freedom of the test
$Q_{obs,t}$ ($m^3 s^{-1}$)	Observed discharge at time step t
$Q_{pred,t}$ ($m^3 s^{-1}$)	Model-predicted discharge at time step t
$\bar{Q}_{obs}$ ($m^3 s^{-1}$)	Mean of observed discharge values over the evaluation period
$\hat{Q}_t$ ($m^3 s^{-1}$)	Predicted discharge at time t after inverse transformation
L_t	Lower bound of the prediction interval on the logarithmic scale
U_t	Upper bound of the prediction interval on the logarithmic scale
MSE_t	Forecast error variance at time step t on the logarithmic scale
$Q_{L,t}$ ($m^3 s^{-1}$)	Lower bound of the prediction

Kang, Y., Xu, Y., Wu, W., Liu, T., Zhang, X., Wang, G. and Quan, L. (2025). Hybrid STL-SARIMA forecasting of reservoir inflows in climate-vulnerable basins: A case study in the Yalong River. *Frontiers in Water* 7, 1674573, https://doi.org/10.3389/frwa.2025.1674573.

Kenyi, M.G.S. and Yamamoto, K. (2024). A hybrid SARIMA-Prophet model for predicting historical streamflow time-series of the Sobat River in South Sudan. *Discover Applied Sciences* 6, 457, https://doi.org/10.1007/s42452-024-06083-x.

Kumar, V., Kedam, N., Sharma, K.V., Mehta, D.J. and Caloiero, T. (2023). Advanced machine learning techniques to improve hydrological prediction: A comparative analysis of streamflow prediction models. *Water* 15(14), 2572, https://doi.org/10.3390/w15142572.

Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models. *Biometrika* 65(2), 297–303, https://doi.org/10.1093/biomet/65.2.297 .

Martinho, A.D., Hippert, H.S. and Goliatt, L. (2023). Short-term streamflow modeling using data-intelligence evolutionary machine learning models. *Scientific Reports* 13, 13824, https://doi.org/10.1038/s41598-023-41113-5.

Nakhaei, M., Zanjanian, H., Nakhaei, P., Gheibi, M., Moezzi, R., Behzadian, K. and Campos, L.C. (2024). Comparative evaluation of deep learning techniques in streamflow monthly prediction of the Zarrine River Basin. *Water* 16(2), 208, https://doi.org/10.3390/w16020208.

Nash, J.E. and Sutcliffe, J.V. (1970). River flow forecasting through conceptual models part I—A discussion of principles. *Journal of Hydrology* 10(3), 282–290, https://doi.org/10.1016/0022-1694(70)90255-6.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* 6(2), 461–464, https://doi.org/10.1214/aos/1176344136.

Sheikholeslami, Z., Ehteshami, M. and Ghasemi, Z. (2024). SARIMA-based prediction of Chalous River flow rates. *Contributions of Science and Technology for Engineering* 1(3), 1–9, https://doi.org/10.22080/cste.2024.27633.1001.

Thota, S., Nassar, A., Filali Boubrahimi, S., Hamdi, S.M. and Hosseinzadeh, P. (2024). Enhancing monthly streamflow prediction using meteorological factors and machine learning models in the Upper Colorado River Basin. *Hydrology* 11(5), 66, https://doi.org/10.3390/hydrology11050066 .

U.S. Geological Survey. *Monitoring location Salmon River at Salmon ID—USGS-13302500*. Water Data for the Nation; https://waterdata .usgs.gov/monitoringlocation/USGS-1330250 https://doi.org/10.1109/TAC.1974.1100705.

Akkala, A., Boubrahimi, S.F., Hamdi, S.M., Hosseinzadeh, P. and Nassar, A. (2025). Improved streamflow forecasting through SWE-augmented spatio-temporal graph neural networks. *Hydrology* 12(10), 268, https://doi.org/10.3390/hydrology12100268.

Alonso Brito, G.R., Rivero Villaverde, A., Lau Quan, A. and Ruíz Pérez, M.E. (2021). Comparison between SARIMA and Holt–Winters models for forecasting monthly streamflow in the western region of Cuba. *SN Applied Sciences* 3, 671, https://doi.org/10.1007/s42452-021-04667-5.

Box, G.E.P., Jenkins, G.M., Reinsel, G.C. and Ljung, G.M. (2015). *Time series analysis: Forecasting and control* (5th ed.). John Wiley & Sons, Hoboken, New Jersey, 712P.

Chu, H., Wang, Z. and Nie, C. (2024). Monthly streamflow prediction of the source region of the Yellow River based on long short-term memory considering different lagged months. *Water* 16(4), 593, https://doi.org/10.3390/w16040593.

Costa, G.E.d.M.e., Menezes Filho, F.C.M.d., Canales, F.A., Fava, M.C., Brandão, A.R.A. and de Paes, R.P. (2023). Assessment of time series models for mean discharge modeling and forecasting in a sub-basin of the Paranaíba River, Brazil. *Hydrology* 10(11), 208, https://doi.org/10.3390/hydrology10110208.

Danandeh Mehr, A., Ghadimi, S., Marttila, H. and Torabi Haghighi, A. (2022). A new evolutionary time series model for streamflow forecasting in boreal lake-river systems. *Theoretical and Applied Climatology* 148, 255–268, https://doi.org/10.1007/s00704-022-03939-3.

Di Nunno, F., de Marinis, G. and Granata, F. (2023). Short-term forecasts of streamflow in the UK based on a novel hybrid artificial intelligence algorithm. *Scientific Reports* 13, 7036, https://doi.org/10.1038/s41598-023-34316-3.

Dickey, D.A. and Fuller, W.A. (1981). Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* 49(4), 1057–1072, https://doi.org/10.2307/1912517.

Islam, K.I., Elias, E., Carroll, K.C. and Brown, C. (2023). Exploring random forest machine learning and remote sensing data for streamflow prediction: An alternative approach to a process-based hydrologic modeling in a snowmelt-driven watershed. *Remote Sensing* 15(16), 3999, https://doi.org/10.3390/rs15163999.

Kamal, S., Wang, J. and Dewan, M.A.A. (2024). Predicting stream flows and dynamics of the Athabasca River Basin using machine learning. *Water* 16(23), 3488, https://doi.org/10.3390/w16233488.