



ارزیابی بهبود ایستایی و دقت پیش‌بینی مدل‌های سری زمانی دبی رودخانه تحت تأثیر رهیافت تفاضل‌گیری (مطالعه موردی: رودخانه دز)

محسن مسلم زاده^۱، سعید فرزین^{۲*}، حجت کرمی^۳

تاریخ ارسال: ۱۳۹۷/۰۲/۲۵

تاریخ پذیرش: ۱۳۹۷/۰۵/۳۰

مقاله پژوهشی

چکیده

هدف از پژوهش حاضر، بررسی تأثیر تفاضل‌گیری‌های فصلی، غیرفصلی و ترکیبی بر میزان ایستایی سری زمانی دبی ماهانه رودخانه دز طی ۶۰ سال می‌باشد. همچنین تأثیر میزان ایستایی بر عملکرد مدل‌های SARIMA در پیش‌بینی سری زمانی از جنبه‌های مختلف بررسی شده است. ابتدا پایداری واریانس سری زمانی بررسی شده و تبدیلات لازم جهت مانا کردن واریانس اعمال گردیده است. سپس به کمک آزمون من-کندال فصلی، ایستایی میانگین ارزیابی شده و از نتایج آن برای ارزیابی وجود روند و نیاز به تفاضل‌گیری استفاده شده است. با استفاده از تفاضل‌گیری‌های فصلی، غیرفصلی و ترکیبی، سه سری جدید ایجاد شده که ایستا بودن آن‌ها به همراه سری اصلی، با تحلیل نمودار ACF و آزمون دیکی-فولر تعمیم یافته مورد بررسی قرار گرفته است. در ادامه، نوع و تعداد پارامترهای موردنیاز در مدل‌ها برای هر کدام از حالات تعیین شده است. نتایج حاکی از آن است که هنگام استفاده از تفاضل‌گیری ترکیبی، تعداد مدل‌های مورد نیاز برای بررسی به شدت کاهش یافته به طوریکه برای مدل‌سازی سری با تفاضل‌گیری فصلی، ۱۹۶ و برای مدل‌سازی سری با تفاضل ترکیبی، تنها ۱۶ مدل نیاز به بررسی داشتند. درحالی‌که نتایج خروجی بهترین مدل از هر دو نوع سری تفاضلی تقریباً نزدیک به هم بود اما بهترین مدل از سری تفاضلی فصلی، با معیارهای ارزیابی $MAE=92/47$ ، $RMSE=154/4$ و $R=0/61$ ، برتری خود را نسبت به دیگر مدل‌ها نشان داد. در نهایت، از مدل منتخب یعنی $SARIMA(4,0,4)(1,1,1)_{12}$ برای پیش‌بینی دبی رودخانه در ۲۴ ماه آینده استفاده شده است.

واژه‌های کلیدی: تفاضل‌گیری، ایستایی، پیش‌بینی، دبی رودخانه، SARIMA

^۱ دانشجوی دکتری، گرایش مهندسی و مدیریت منابع آب، دانشکده مهندسی عمران، دانشگاه سمنان، ۰۹۱۶۶۴۱۴۰۱۴، moslemzadeh.m@semnan.ac.ir

^۲ استادیار، گروه مهندسی آب و سازه‌های هیدرولیکی، دانشکده مهندسی عمران، دانشگاه سمنان، ۰۹۱۷۴۱۱۲۸۹۱، saeed.farzin@semnan.ac.ir (نویسنده مسئول)

^۳ استادیار، گروه مهندسی آب و سازه‌های هیدرولیکی، دانشکده مهندسی عمران، دانشگاه سمنان، hkarami@semnan.ac.ir، ۰۹۱۲۴۸۰۳۳۵۰

مقدمه

امروزه با توجه به رشد بی‌رویه جمعیت جهان و به دنبال آن افزایش مصرف آب، برنامه‌ریزی و مدیریت منابع آب به یکی از شاخصه‌های مهم پیشرفت هر کشور تبدیل شده است. تأمین و برنامه‌ریزی تخصیص آب در بخش‌های مختلف کشاورزی، فضای سبز، صنعت، شرب و بهداشت نیازمند داشتن اطلاعات پایه از وضعیت موجود منابع آب و همچنین پیش‌بینی آن‌ها در آینده است. در این راستا، مدل‌سازی و پیش‌بینی سری‌های زمانی هیدرولوژیکی از اهمیت بالایی برخوردار است. بدین منظور، تاکنون مدل‌های مختلفی ارائه شده است که از جمله آن‌ها می‌توان مدل‌های خود همبسته با میانگین متحرک (ARMA)^۴، خود همبسته با میانگین متحرک تجمعی (ARIMA)^۵ و خود همبسته با میانگین متحرک تجمعی فصلی (SARIMA)^۶ را نام برد.

توسط مدل‌های مذکور، مطالعات هیدرولوژیکی در زمینه‌های متنوعی مانند پیش‌بینی میزان بارش (Dabral et al., 2017)، مواد جامد محلول در جریان‌های سطحی (Salmani and Jajaei, 2016)، پارامترهای کیفی آب رودخانه (حقی زاده و همکاران، ۱۳۹۶)، پارامترهای کیفی دمای هوا (Afrifa-Yamoah, 2015)، سطح آب مخازن سد (Yang et al., 2017)، تبخیر (Bazrafshan et al., 2017)، خشکسالی (Bazrafshan et al., 2015) و سطح آب زیرزمینی (Patle et al., 2015) صورت گرفته است. یکی از مهم‌ترین متغیرهای هیدرولوژیکی، دبی رودخانه‌ها است که همواره مورد استقبال محققان قرار گرفته است. (Al-Juboori and Guven, 2016) با ارائه مدلی تلفیقی با برنامه ژنتیک اقدام به پیش‌بینی جریان رودخانه کرده و سپس نتایج آن را با سری زمانی مقایسه کردند. نتایج حاکی از برتری مدل ارائه شده داشت. (Elganiny and Eldwer, 2016) به بررسی و

مقایسه مدل خود همبسته با میانگین متحرک و مدل خود همبسته با میانگین متحرک تجمعی فصلی برای تعدادی از رودخانه‌ها پرداختند و نشان دادند که مدل غیر فصلی نتایج بهتری دارد. سیدیان و همکاران، (۱۳۹۳) دقت روش‌های سری زمانی و ماشین پشتیبان SVM^۷ جهت پیش‌بینی دبی جریان رودخانه را مقایسه نمودند. نشان دادند که دقت روش SVM بالاتر و در هر دو روش دقت در مقیاس ماهانه بالاتر از روزانه است.

یکی از اساسی‌ترین فرضیات در مدل‌سازی سری زمانی، مانا بودن و ایستایی سری زمانی است (Brockwell and Davis, 1991). نا ایستا بودن سری زمانی می‌تواند در میانگین و واریانس اثر خود را جلوه‌گر کند. چنانچه یک سری زمانی نا ایستا باشد روش‌هایی وجود دارد که می‌توان آن را ایستا نمود. از جمله آن‌ها می‌توان شناسایی و حذف مستقیم عوامل نا ایستایی (Moeeni et al., 2017) و استفاده از توابع تبدیل و تفاضل‌گیری را نام برد. از کاربردهای تفاضل‌گیری، حذف کامل یا کاهش اثر عوامل نا ایستایی مثل مؤلفه‌های روند و تناوب است. این مؤلفه‌ها را می‌توان به ترتیب با تفاضل‌گیری‌های غیر فصلی و فصلی از بین برد (Pena et al., 2017). (خادمی و همکاران، ۱۳۹۶) نشان دادند که در سری زمانی تراز دریاچه میشیگان-هارن، تفاضل‌گیری توأم باعث ایستا سازی هر چه بیشتر و همچنین بهبود نتایج پیش‌بینی می‌گردد.

در پژوهش حاضر به منظور بررسی تأثیر عملکرد تفاضل‌گیری در بهبود ایستایی و نتایج مدل‌سازی سری زمانی دبی رودخانه، انواع تفاضل‌گیری‌های فصلی، غیر فصلی و توأم بر روی داده‌های مربوط به متوسط دبی جریان ماهانه رودخانه دز در بازه زمانی ۶۰ ساله (۱۳۳۶-۱۳۹۵) اعمال می‌گردد. ابتدا عملکرد آن‌ها در ایستا سازی سری زمانی مورد تجزیه و تحلیل قرار می‌گیرد. سپس تأثیر این تفاضل‌گیری‌ها بر تعداد مدل‌های مورد نیاز نیز بررسی خواهد شد. در نهایت،

⁷ Support Vector Machine

⁴ Auto Regressive Moving Average

⁵ Auto Regressive Integrated Moving Average

⁶ Seasonal ARIMA model

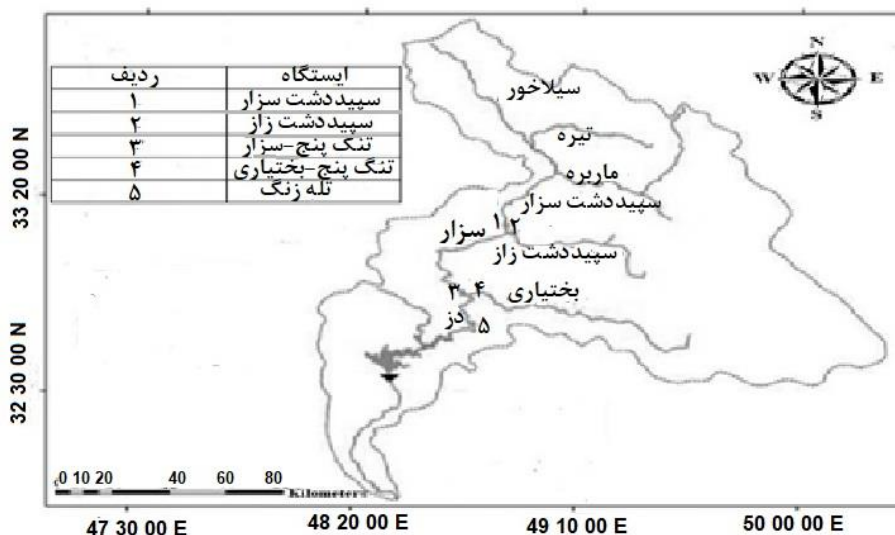
شمال به جنوب است. همان‌گونه که در شکل (۱) مشخص است در سرچشمه‌ها و در مسیر رودخانه دز چندین ایستگاه هیدرومتری وجود دارد که در تحقیق حاضر به دلیل موقعیت جغرافیایی و همچنین کامل‌تر بودن اطلاعات اندازه‌گیری شده، از داده‌های مربوط به دبی متوسط ماهانه از سال آبی ۱۳۳۶ تا سال ۱۳۹۵ ثبت شده در پنجمین ایستگاه هیدرومتری یعنی ایستگاه تله زنگ استفاده گردیده است. از ۷۲۰ داده موجود، ۵۷۶ داده ابتدایی یعنی معادل ۸۰ درصد داده‌ها برای واسنجی و ۱۴۴ داده انتهایی برای اعتبار سنجی استفاده شده است. شکل (۲) نمودار سری زمانی دبی ماهانه رودخانه دز و جدول (۱) مشخصه‌های آماری این سری را نشان می‌دهند.

ضمن مقایسه دقت پیش‌بینی مدل‌های مختلف و انتخاب بهترین مدل، از مدل منتخب برای پیش‌بینی ۲۴ ماه آینده استفاده شده است.

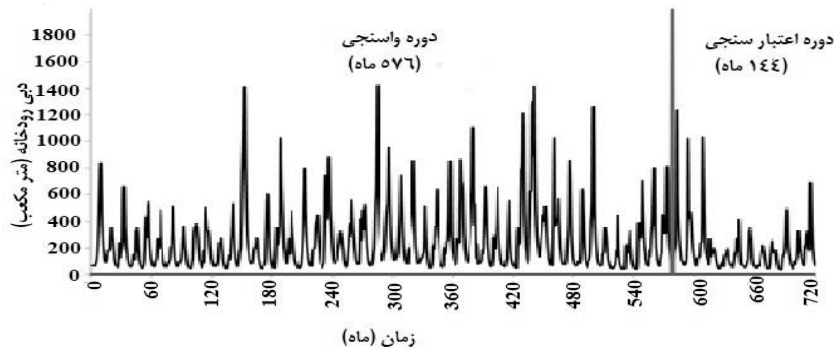
مواد و روش‌ها

داده‌ها و منطقه مورد مطالعه

شکل (۱) موقعیت حوزه آبخیز دز در کشور ایران را نشان می‌دهد. این حوزه از لحاظ موقعیت جغرافیایی بین ۴۸ درجه و ۱۰ دقیقه تا ۵۰ درجه و ۲۱ دقیقه طول شرقی و ۳۱ درجه و ۳۴ دقیقه تا ۳۴ درجه و ۷ دقیقه عرض شمالی قرار دارد. مساحت کل حوزه بالغ بر ۲۱۷۲۰ کیلومترمربع و متوسط ارتفاع آن حدود ۱۶۰۰ متر می‌باشد. جهت جریان در حوضه آبریز دز از سمت



شکل (۱): موقعیت حوزه آبخیز دز و ایستگاه‌های هیدرومتری واقع در مسیر رودخانه دز.



شکل (۲): سری زمانی دبی ماهانه رودخانه دز.

جدول (۱): مشخصه‌های آماری سری زمانی دبی رودخانه دز

مشخصه	تعداد	میانگین	انحراف	واریانس	حداقل	حداکثر	ضریب	ضریب
آماري	داده‌ها		معیار		داده‌ها	داده‌ها	چولگی	کشیدگی
دبی رودخانه (مترمکعب بر ثانیه)	۷۲۰	۲۳۷/۹۶	۲۲۲/۰۸	۴۹۳۲۱/۲۶	۳۶/۲۶	۱۴۲۱	۲/۲۳	۶/۱۷

که در روابط مذکور، $x(t)$ سری زمانی، E و Var توابع امید ریاضی و واریانس، γ_s تابع اتوکواریانس سری زمانی، m_1 و m_2 اعداد ثابت می‌باشند.

مدل‌های احتمالاتی سری زمانی، برای سری ایستا در واریانس و میانگین تعریف شده‌اند. بنابراین ایستایی سری زمانی ابتدا در واریانس و سپس در میانگین بررسی می‌شود. در این مطالعه، برای بررسی ایستایی واریانس از روش کاکس-باکس^۸ و برای ایستا سازی میانگین از رهیافت تفاضل‌گیری های غیر فصلی و فصلی با گام زمانی معادل دوره تناوب فصلی (ω) استفاده شده است.

در روش کاکس-باکس چنانچه سری زمانی فاقد ایستایی واریانس باشد، تبدیل کاکس-باکس به صورت رابطه (۵) بر روی داده‌های سری زمانی تعریف می‌شود:

$$T(x_t) = \frac{x_t^\lambda - 1}{\lambda} \quad (5)$$

ایستایی سری زمانی

اگر در یک سری زمانی میانگین، واریانس و همچنین کوواریانس در دوره‌های مشخص ثابت باشد (رابطه‌های ۱ تا ۴)، آن سری ایستا و چنانچه برای نقاط زمانی t_1 تا t_k در هر تأخیر زمانی h ، توزیع توأم $x(t_1)$ تا $x(t_n)$ مانند توزیع توأم $x(t_1+h)$ و $x(t_n+h)$ باشد، سری زمانی ایستای اکید خواهد بود (Salas et al., 1980). به عبارت دیگر مفهوم اساسی ایستایی این است که قوانین احتمالی حاکم بر فرآیند با زمان تغییر نکرده و فرآیند در تعادل آماری باقی می‌ماند.

$$E(|x(t)|^2) < \infty, t \in Z \quad (1)$$

$$E(x(t)) = m_1, t \in Z \quad (2)$$

$$Var(x(t)) = m_2, t \in Z \quad (3)$$

$$\gamma_s(r, s) = \gamma_s(r+t, s+t), \{r, s, t\} \in Z \quad (4)$$

فرض کرد که $cov(S_g, S_h) = 0$. در نهایت، آماره‌ای با عنوان Z' از رابطه (۹) به دست می‌آید:

$$Z' = \left\{ \begin{array}{l} \frac{S' - 1}{\sigma_{S'}}, S' > 0 \\ 0, S' = 0 \\ \frac{S' + 1}{\sigma_{S'}}, S' < 0 \end{array} \right. \quad (9)$$

چنانچه مقدار سطح احتمال α که با توجه به مقدار Z' محاسبه می‌گردد کمتر از ۵ درصد باشد نشان‌دهنده آن است که در سری زمانی روند وجود دارد (باقرپور و همکاران، ۱۳۹۶). بعد از انجام این آزمون و در صورت وجود روند، اقدام به حذف روند به کمک روش‌هایی مانند تفاضل گیری شده و سپس مدل مناسب برازش داده می‌شود.

کنترل مجدد ایستایی

از آنجاکه ایستا بودن سری زمانی یک فرض پایه و اساسی در مدل‌سازی و پیش‌بینی آن است، جهت حصول اطمینان از اینکه بعد از تفاضل گیری‌های غیر فصلی، فصلی و ترکیبی، سری‌های مذکور حتماً ایستا شده‌اند، می‌توان با روش‌های مختلفی این موضوع را کنترل نمود (خادمی و همکاران، ۱۳۹۶). بدین منظور در این پژوهش از آزمون ایستایی دیکی فولر تعمیم یافته^{۱۱} استفاده گردیده است. این آزمون (روابط ۱۰ تا ۱۳) بر اساس وجود یا عدم وجود ریشه واحد عمل کرده به گونه‌ای که با بررسی عدم وجود ریشه واحد، ایستا بودن سری را مشخص می‌کند.

$$\Delta x(t) = \alpha + \beta_1 t + \beta_2 t^2 + \gamma x(t-1) \quad (10)$$

$$+ \phi_1 \Delta x(t-1) + \dots + \phi_{p-1} \Delta x(t-p+1) + \varepsilon(t)$$

$$\tau = \frac{\hat{\gamma}}{\hat{\sigma}_{\gamma}} \quad (11)$$

$$H_0: \gamma = 0 \quad (12)$$

که در آن λ پارامتر تبدیل، x_t مقدار متغیر اولیه در زمان t و $T(x_t)$ مقدار تبدیل یافته نظیر می‌باشد. تبدیل‌های کاکس-باکس صرفاً در جهت پایداری واریانس عمل نکرده بلکه تقریب بهتری برای نرمال بودن نیز ایجاد می‌کند. (Salas et al., 1980)

بیشتر پارامترهای اقلیمی و هیدرولوژیکی به صورت فصلی هستند که نوسانات فصلی باعث پیچیدگی بررسی روند می‌شود. در خیلی از رودخانه‌ها قسمتی از جریان ناشی از باران و بخشی ناشی از ذوب برف است. همچنین در این رودخانه‌ها، افزایش دما طی اوقات مختلف سال باعث تغییر در میزان رواناب در فصل بارش و ذوب برف می‌شود. هنگامی که داده‌های هر فصل به صورت جداگانه بررسی گردند، این روندها به خوبی قابل مشاهده می‌باشند. بنابراین نیاز است تا داده‌ها به صورت فصلی هم مورد بررسی قرار گیرد [۱۷]. در این پژوهش از آزمون من-کندال فصلی^۹ (رابطه‌های ۶ تا ۹) برای تعیین روند در داده‌ها استفاده شده است.

$$S' = \sum_{j=1}^p S_j \quad (6)$$

که در آن P تعداد کل فصول و S_j آماره کندال مربوط به فصل j ام ($j=1, 2, \dots, p$) بوده و طبق رابطه (۷) محاسبه می‌شود:

$$S_j = \sum_{i=1}^{n_j} \sum_{k=i+1}^{n_j-1} \text{sgn}(x(i) - x(k)) \quad (7)$$

اگر در داده‌های سری زمانی همبستگی متوالی وجود داشته باشد، واریانس از رابطه (۸) قابل محاسبه است. در صورت عدم وجود همبستگی متوالی در داده‌ها، ترم دوم سمت راست رابطه حذف می‌شود.

$$\sigma_{S'}^2 = \sum_{j=1}^p \text{Var}(S_j) + \sum_{g=1}^{p-1} \sum_{h=g+1}^p \sigma_{gh} \quad (8)$$

که σ_{gh} کوواریانس بین آماره کندال در فصل g و فصل h را نشان می‌دهد. با فرض استقلال داده‌ها می‌توان

¹¹ ADF

9 Seasonal Mann-Kendall
 10 Probability Value

مدل‌های سری زمانی

به منظور پیش‌بینی دبی رودخانه و با توجه به وجود مدل‌های مختلف سری زمانی بر اساس نظریه‌های باکس-جنکینز^{۱۳}، در این پژوهش از مدل‌های آریمای فصلی مکرر^{۱۴} استفاده گردیده است. این مدل با روابط زیر مطرح می‌شود:

$$SARIMA(p,d,q)(P,D,Q): \quad (15)$$

$$\Phi(B^w)\varphi(B)(1-B^w)^D(1-B)^d x(t)$$

$$= \Theta(B^w)\theta(B)\varepsilon(t)$$

$$\Phi(B^w) = (1 - \Phi_1 B^w - \Phi_2 B^{2w} - \dots - \Phi_p B^{pw}) \quad (16)$$

$$\varphi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p) \quad (17)$$

$$\Theta(B^w) = (1 - \Theta_1 B^w - \Theta_2 B^{2w} - \dots - \Theta_q B^{qw}) \quad (18)$$

$$\theta(B) = (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) \quad (19)$$

که در آن $x(t)$ سری زمانی دبی رودخانه، $\varepsilon(t)$ سری باقی‌مانده‌ها، p مرتبه پارامتر خود همبسته غیر فصلی، q مرتبه پارامتر میانگین متحرک غیر فصلی، P مرتبه پارامتر خود همبسته فصلی، Q مرتبه پارامتر میانگین متحرک فصلی، w دوره تناوب، φ پارامتر خود همبسته غیر فصلی، θ پارامتر میانگین متحرک غیر فصلی، Φ پارامتر خودهمبستگی فصلی، Θ پارامتر میانگین متحرک فصلی، B عملگر تفاضلی، $(1-B^w)^D$ برابر D امین تفاضل‌گیری فصلی با گام w و $(1-B)^d$ برابر d امین تفاضل‌گیری غیر فصلی می‌باشد.

حداکثر مرتبه پارامترهای مدل‌های مذکور (p, q, P, Q) را می‌توان از روی نمودار ACF شناسایی کرد. اگر مرتبه‌های موردنیاز برای پارامترهای هرمدل به صورت: $p = \{0, 1, \dots, l\}$ ، $q = \{0, 1, \dots, m\}$ ، $P = \{0, 1, \dots, n\}$ و $Q = \{0, 1, \dots, o\}$ باشند آنگاه تعداد مدل‌های موردنیاز (NOM) با توجه به قانون احتمالاتی و بر اساس رابطه (۲۰) به دست می‌آید:

$$H_1: \gamma < 0 \quad (13)$$

که در آن Δ عملگر تفاضلی مرتبه اول، α ، β_1 و β_2 به ترتیب مقدار ثابت و ضرایب روند خطی و درجه‌دو، φ_i ضریب i امین مرتبه خودهمبستگی، p حداکثر مرتبه خودهمبستگی در نظر گرفته شده، τ آماره آزمون دیکي-فولر تعمیم‌یافته، γ ضریب یا ریشه تخمین زده شده، σ_γ خطای استاندارد تخمین زده شده، H_0 فرض صفر مبنی بر وجود ریشه واحد و H_1 فرض جایگزین مبنی بر عدم وجود ریشه واحد است. با محاسبه مقدار τ در صورتی که این آماره از مقدار بحرانی در توزیع دیکي-فولر کوچک‌تر باشد فرض صفر رد می‌شود که به معنی آن است که ریشه واحدی وجود نداشته و سری ایستا خواهد بود.

با عنایت به اینکه ضعف آزمون دیکي-فولر تعمیم‌یافته در این است که اثر مؤلفه تناوبی بر نا ایستاد شدن سری را در نظر نمی‌گیرد، بنابراین برای سری‌های فصلی استفاده از دیگر آزمون‌ها در کنار آن نیز ضروری خواهد بود. یکی از روش‌های آزمون مجدد ایستایی، رسم نمودار همبستگی نگار و بررسی شهودی آن است. این نمودار مقادیر تابع خودهمبستگی (ACF)^{۱۲} را به ازای گام‌های زمانی مختلف نشان می‌دهد. در صورتی که سری زمانی ایستا باشد، این نمودار بلافاصله به صورت نمایی یا نوسانی به صفر میرا خواهد شد (Cryer and Chan, 2008). در این نمودار مقدار ضریب خودهمبستگی داده‌ها (ρ) برای گام‌های مختلف رسم خواهد شد. مقدار ضریب خودهمبستگی برای گام h ام به صورت رابطه (۱۴) به دست می‌آید:

$$\rho(h) = \frac{\sum_{t=h}^n (x(t) - \bar{x})(x(t-h) - \bar{x})}{\sum_{t=h}^n (x(t) - \bar{x})^2} \quad (14)$$

$$h = 1, 2, \dots, n/4$$

¹⁴ Multiplicative Seasonal ARIMA

¹² Auto Correlation Function

¹³ BOX-JENKINS

ترسیم و مورد تفسیر قرار گرفته که در اشکال (۶) الی (۸) آورده شده است.

علاوه بر روش‌های نموداری، از آزمون پرت-مانتو^{۱۶} نیز برای بررسی مناسب بودن مدل استفاده شده است. این آزمون از خودهمبستگی باقیمانده‌ها برای بررسی فرضیه صفر (H_0) یعنی عدم وجود همبستگی بین باقیمانده‌های مدل به کمک آماره آزمون Q استفاده می‌کند. هرگاه مقدار آماره Q از مقدار متناظر جدول احتمالاتی کی دو^{۱۷} (χ^2) بیشتر باشد فرضیه (H_0) که گاهی آن را فرضیه کفایت مدل نیز می‌نامند رد می‌شود. در این پژوهش مقادیر احتمالاتی متناظر با این آزمون (P-Value) برای ۹۵ درصد اطمینان لحاظ و استقلال باقیمانده‌ها مورد ارزیابی قرار گرفته است.

نتایج و بحث

قبل از مدل‌سازی سری زمانی دبی رودخانه دز و به‌منظور بررسی ایستا بودن سری، ابتدا ایستا بودن سری از نظر واریانس بررسی گردیده است. برای این کار ابتدا پارامتر تبدیل توانی باکس-کاکس برای سری موردنظر محاسبه شد. با توجه به ضرایب به‌دست‌آمده، دومرتبه تبدیل انجام پذیرفته تا به حالت مطلوب رسیده شود. با نخستین تبدیل ($W1_t = \frac{1}{\sqrt{x(t)}}$) سری تبدیل یافته اول ($W1_t$) و با اعمال تبدیلی دیگر، دومین سری تبدیل یافته ($W2_t = \sqrt{W1_t} = \frac{1}{\sqrt[4]{x(t)}}$) به‌دست‌آمده که این سری از نظر واریانس حالت ایستا پیدا کرد. نتایج پارامتر تبدیل توانی سری اولیه $x(t)$ و سری تبدیل یافته دوم $W2_t$ جهت بررسی ایستایی واریانس در اشکال (۳ و ۴) آورده شده است.

$$NOM_{SARIMA} = (l+1)(m+1)(n+1)(o+1) \quad (20)$$

معیارهای ارزیابی مدل‌ها

به‌منظور دستیابی به دقیق‌ترین مدل، معیارهای ارزیابی متداول معتبر بر روی سری زمانی مورد مطالعه، برازش داده می‌شود. در اینجا از معیارهای ضریب همبستگی (R)، جذر میانگین مربعات خطاها ($RMSE$) و میانگین مطلق خطاها (MAE) برای ارزیابی و تحلیل نتایج مدل‌ها استفاده گردیده است. هرچه مقادیر $RMSE$ و MAE به صفر نزدیک‌تر باشند نشان‌دهنده عملکرد بهتر مدل است. مقدار ضریب همبستگی که بین -۱ و ۱ متغیر است در حالت برتر به مقدار صفر نزدیک‌تر می‌باشد.

پس از بررسی معیارهای ارزیابی و انتخاب بهترین مدل، مناسب بودن باقیمانده‌های حاصل از نتایج آن مورد بررسی قرار گرفته و در صورت تأیید، مدل برگزیده انتخاب خواهد شد. در انتها با استفاده از مدل منتخب، پارامتر دبی رودخانه در آینده پیش‌بینی می‌شود.

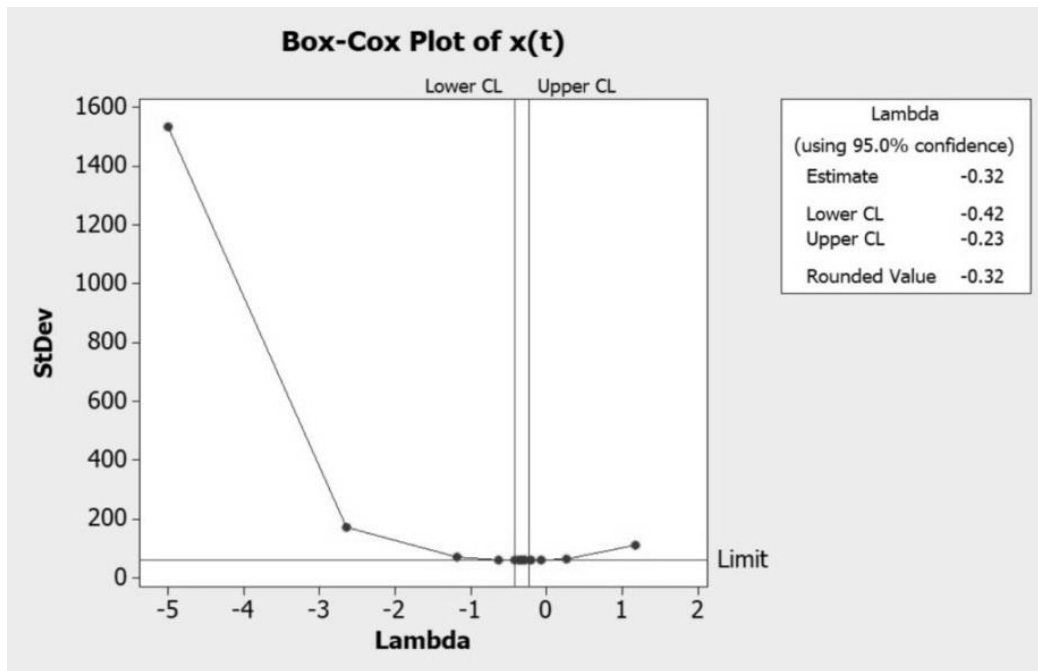
بررسی استقلال باقی‌مانده‌ها

یکی از شروط اصلی برای تأیید صحت مدل انتخابی سری زمانی آن است که باقیمانده‌های حاصل از برازش مدل باید تقریباً دارای خواص متغیرهای تصادفی نرمال مستقل هم توزیع با میانگین صفر و واریانس ثابت باشند. در این پژوهش برای اطمینان از صحت مدل برازش شده، باقیمانده مدل برازش شده مورد بررسی و تجزیه تحلیل قرار گرفت. برای این کار نمودار احتمال نرمال باقیمانده‌ها، نمودار باقیمانده‌ها در مقابل مقادیر برازش داده‌شده، نمودار باقیمانده‌ها در طول زمان، هیستوگرام باقیمانده‌ها و نمودارهای توابع خودهمبستگی و خودهمبستگی جزئی^{۱۵} باقیمانده‌ها

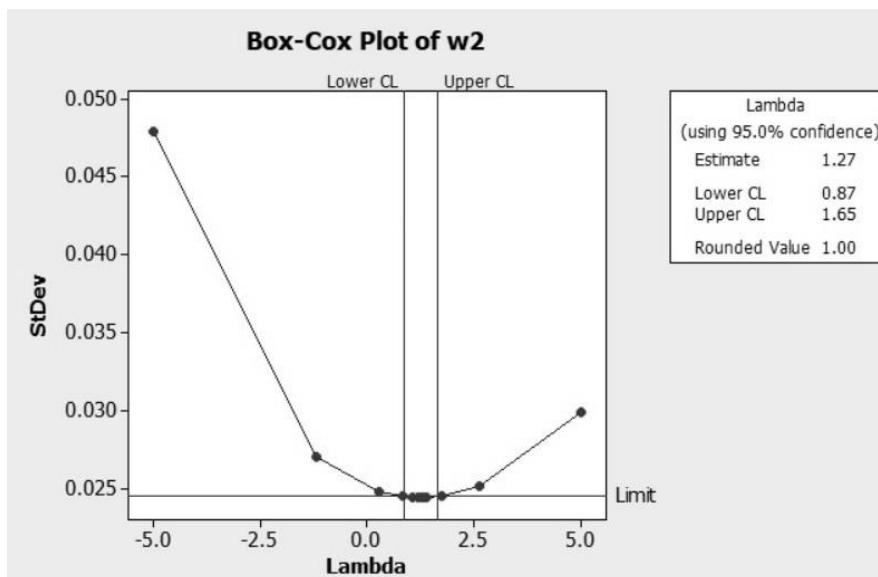
¹⁷ Chi Square

¹⁵ PACF

¹⁶ portmanteau test



شکل (۳): نتایج پارامتر تبدیل توانی برای سری زمانی اولیه دبی رودخانه دز ($x(t)$)



شکل (۴): نتایج پارامتر تبدیل توانی برای سری زمانی تبدیل یافته دوم ($W2_t$)

ارزیابی قرار گرفته و در نهایت پس از مدل سازی و حصول نتایج، اعداد به حالت اولیه تبدیل معکوس شدند.

بهمنظور بررسی وجود روند و ایستایی از نظر میانگین در سری فوق، از آزمون من-کندال فصلی

با توجه به مقدار به دست آمده برای ضریب تبدیل کاکس-باکس، مشخص می شود که سری تبدیل یافته $W2_t$ از نظر واریانس ایستا شده و نیاز به تبدیل دیگری ندارد. سری زمانی تبدیل یافته دوم تا انتها مورد

وجود روند در این سری محرز می‌باشد که باید از روش-
هایی همچون تفاضل گیری اقدام به حذف روند نمود.

استفاده گردید. نتایج این آزمون در جدول (۲) نشان
داده شده است. با توجه به مقدار آماره احتمال آزمون
(P-Value) که از سطح معناداری ۵ درصد کمتر است،

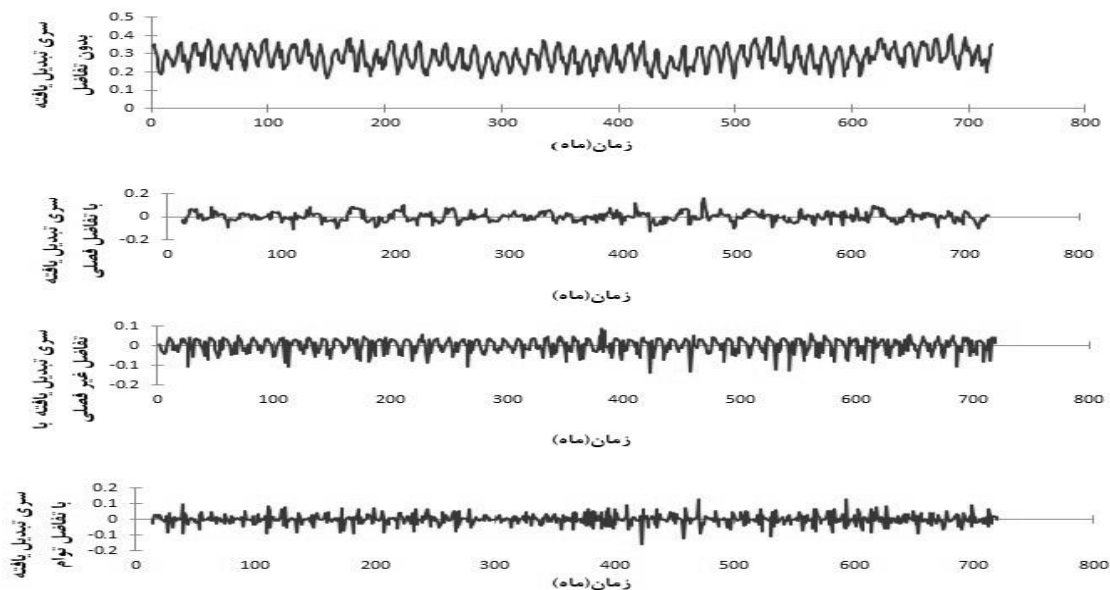
جدول (۲): نتایج آزمون من کندال فصلی برای سری تبدیل یافته

P-Value	S'	τ	انحراف معیار	حداکثر (m^3/s)	میانگین (m^3/s)	حداقل (m^3/s)	تعداد مشاهده
۰/۰۴۵	۱۰۸۷	۰/۰۵۱	۰/۰۵۴	۰/۴۰۸	۰/۲۸۲	۰/۱۶۳	۷۲۰

($D=1$)، می‌توان این عمل را به همراه تفاضل گیری
غیر فصلی مرتبه اول ($d=1$) و تفاضل گیری توأم (هم
غیر فصلی و هم فصلی) بر روی داده‌ها انجام داد. بدین
منظور سری اصلی به سه سری جدید تبدیل می‌شود.
شکل (۵) این سری‌های زمانی را نشان داده است.
با بررسی نمودارهای شکل (۵) مشاهده می‌شود که
نمودار سری‌های تفاضل گیری شده کاملاً افقی بوده و
در نتیجه روندی در آن‌ها وجود ندارد.

یکی از راهکارهای حذف روند در سری زمانی
استفاده از عملیات تفاضل گیری است. در هنگام
استفاده از تفاضل گیری فصلی، گام تفاضل گیری باید
مناسب بوده تا نتیجه مطلوب که همان ایستا شدن
سری است، حاصل شود. در اینجا از گام ۱۲ برای تفاضل
گیری فصلی دبی رودخانه استفاده شده است.

با مشخص شدن گام مناسب تفاضل گیری فصلی
($\omega=12$) برای تفاضل گیری فصلی از مرتبه اول



شکل (۵): سری‌های زمانی تفاضل گیری شده

داشته باشند؛ بنابراین برای اطمینان از ایستا بودن
سری‌های اصلی و تفاضل گیری شده، آزمون دیکی-
فولر تعمیم یافته به کار برده شد که نتایج آن در جدول
(۳) آورده شده است.

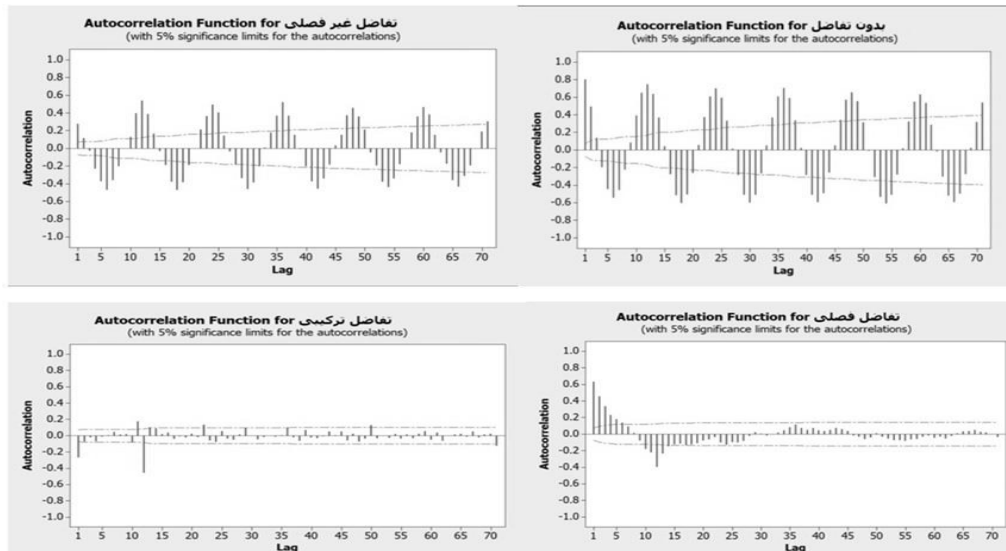
هرچند مشاهده شد که روند به‌عنوان عامل مهم نا
ایستایی در سری‌های تفاضلی دبی رودخانه وجود ندارد،
اما نمی‌توان با اطمینان بیان داشت که سری ایستا است؛
زیرا ممکن است دیگر عوامل نا ایستایی همچنان وجود

جدول (۳): نتایج آزمون ADF برای سری تبدیل یافته

نوع آزمون	حد بحرانی	d=0, D=0	d=1, D=0	d=0, D=1	d=1, D=1
ADF دیکی-فولر	-۳/۳۹۶	(ایستا)	(ایستا)	(ایستا)	(ایستا)
تعمیم یافته	-۶/۰۳۴	-۶/۰۳۴	-۲۱/۴۱۵	-۸/۵۶۶	-۱۰/۱۲۲

با توجه به جدول (۳) مشخص می‌شود که تمامی این سری‌ها از نظر وجود روند، ایستا هستند. همچنین مشاهده می‌شود که مقادیر آماره‌ها برای تفاضل گیری غیر فصلی، توأم، فصلی و سری اصلی به ترتیب اختلاف بیشتری با حدود بحرانی داشته است. از این رو می‌توان دریافت چهار سری مذکور به ترتیب دارای ایستایی بیشتری هستند. آزمون دیکی-فولر برای تعیین ایستایی سری‌های فصلی گاهی ضعیف عمل می‌کند و برای سری مورد بررسی در این پژوهش، در مورد فصلی بودن داده‌های تفاضل گیری شده اطلاعاتی در دسترس نیست. لذا به منظور اطمینان بیشتر، با رسم نمودار همبستگی نگار،

مجدداً ایستا بودن انواع سری‌های بدون تفاضل و با تفاضل گیری‌های غیر فصلی، فصلی و توأم بررسی شده است که نتایج آن در شکل (۶) آورده شده است. با توجه به شکل (۶) مشاهده می‌شود که سری‌های تبدیل یافته با تفاضل گیری توأم و تفاضل گیری فصلی پس از چند گام به صفر نزدیک شده و داخل حدود اطمینان ۹۵٪ قرار گرفته‌اند؛ بنابراین در این مرحله ایستا بودن آن‌ها محرز می‌شود. اما سری تبدیل یافته بدون تفاضل و با تفاضل گیری غیر فصلی، نامیرا و نوسانی بوده که نشان دهنده آن است که سری‌های مذکور نایستا می‌باشند، موضوعی که آزمون دیکی فولر در شناسایی آن ناتوان بود.



شکل (۶): نمودارهای خودهمبستگی حالات مختلف سری زمانی تبدیل یافته

خودهمبستگی مشاهده می‌شود که تفاضل گیری فصلی، تا ۶ گام غیر فصلی و یک گام فصلی با تأخیر ۱۲ دارای خودهمبستگی بوده درحالی‌که در تفاضل گیری توأم (ترکیبی)، این مقادیر به ۱ گام غیر فصلی و یک

از این موضوع می‌توان نتیجه گرفت که حتی تفاضل گیری مرتبه اول همیشه باعث ایستا شدن سری نخواهد شد و همواره بعد از این عمل باید ایستایی مورد بررسی مجدد قرار گیرد. همچنین با توجه به گراف‌های

سری‌های زمانی نیز تعیین می‌گردد. این تعداد که بر اساس گام‌های با مقادیر بزرگ از این نمودار به دست آمد در جدول (۴) ارائه شده است. در این جدول همچنین تعداد مدل‌های موردنیاز برای دستیابی به دقیق‌ترین مدل با توجه به رابطه (۲۰) ارائه شده است.

جدول (۴): مرتبه‌های پارامتر و تعداد مدل‌های موردنیاز

تعداد مدل	Q	P	q	p	نوع سری
سری نایستا					بدون تفاضل
سری نایستا					تفاضل‌گیری غیر فصلی
۱۹۶	۱	۱	۶	۶	تفاضل‌گیری فصلی
۱۶	۱	۱	۱	۱	تفاضل‌گیری توأم

گام فصلی با تأخیر ۱۲ کاهش داده شده است. این قضیه بهبود ایستایی را در سری با تفاضل‌گیری توأم نشان می‌دهد.

از روی همبستگی نگارها علاوه بر ایستایی، بیش‌ترین تعداد پارامترهای موردنیاز برای مدل‌های

ترکیب $SARIMA(p,1,q)(P,1,Q)_{12}$ در نظر گرفته شد. موضوع دیگری که از نمودارهای شکل (۶) و جدول (۴) می‌توان دریافت نوع مدل مناسب برای مدل‌سازی و پیش‌بینی است؛ بنابراین با جایگذاری مرتبه‌های مختلف p, Q, P و q ، سری تبدیل یافته توسط حالات مختلف، مدل‌سازی و پس‌از آن در دوره اعتبارسنجی پیش‌بینی شده است. آنگاه نتایج خروجی که برای سری تبدیل یافته محاسبه شده بودند به صورت معکوس به شکل اعداد اولیه تبدیل شدند. معیارهای ارزیابی برای نتایج تمامی مدل‌ها محاسبه و دقیق‌ترین مدل بر اساس همین معیارها انتخاب گردید.

با توجه به جدول (۴)، مشخص می‌گردد که تعداد مدل‌ها در تفاضل‌گیری توأم به میزان قابل توجهی نسبت به تفاضل‌گیری فصلی کاهش می‌یابد؛ بنابراین ایستایی بیشتر سری به کمک تفاضل‌گیری، این مزیت را هم دارد که با بررسی تعداد کمتری مدل می‌توان به دقیق‌ترین مدل دست یافت. همچنین با توجه به آنکه سری اصلی و سری حاصل از تفاضل‌گیری غیر فصلی نایستا بودند مدل‌سازی برای آن‌ها انجام نمی‌شود. در این پژوهش مدل مناسب برای سری حاصل از تفاضل‌گیری فصلی به صورت $SARIMA(p,0,q)(P,1,Q)_{12}$ و برای سری حاصل از تفاضل‌گیری توأم بصورت

جدول (۵): نتایج ارزیابی دقیق‌ترین مدل‌ها.

مدل	واسنجی			اعتبارسنجی		
	RMS	MA	R	RMS	MA	R
(۱.۱.۱)	۱۲۲/۸۱	۶۷/۷	۰/۸۵	۱۷۰/۱۷	۱۰۷/۷	۰/۶
(۱.۱.۱) _۳						
(۴.۰.۴)	۱۲۱/۸	۶۷/۲	۰/۸۵	۱۵۴/۴	۹۲/۴۷	۰/۶۱
(۱.۱.۱) _۳						

تفاضل فصلی نسبت به سری با تفاضل توأم، بهتر و بالاتر است.

دقیق‌ترین مدل‌ها به همراه نتایج معیارهای ارزیابی آن‌ها چه برای دوره واسنجی و چه برای دوره اعتبارسنجی در جدول (۵) آورده شده است. مشاهده می‌شود که دقت مدل‌ها تا حدی مخصوصاً در قسمت واسنجی نزدیک به هم هستند اما دقت مدل $SARIMA$ با

بررسی کیفیت مدل‌ها

برای اطمینان از صحت مدل انتخاب شده به تجزیه و تحلیل باقیمانده‌های مدل برازش شده پرداخته می‌شود.

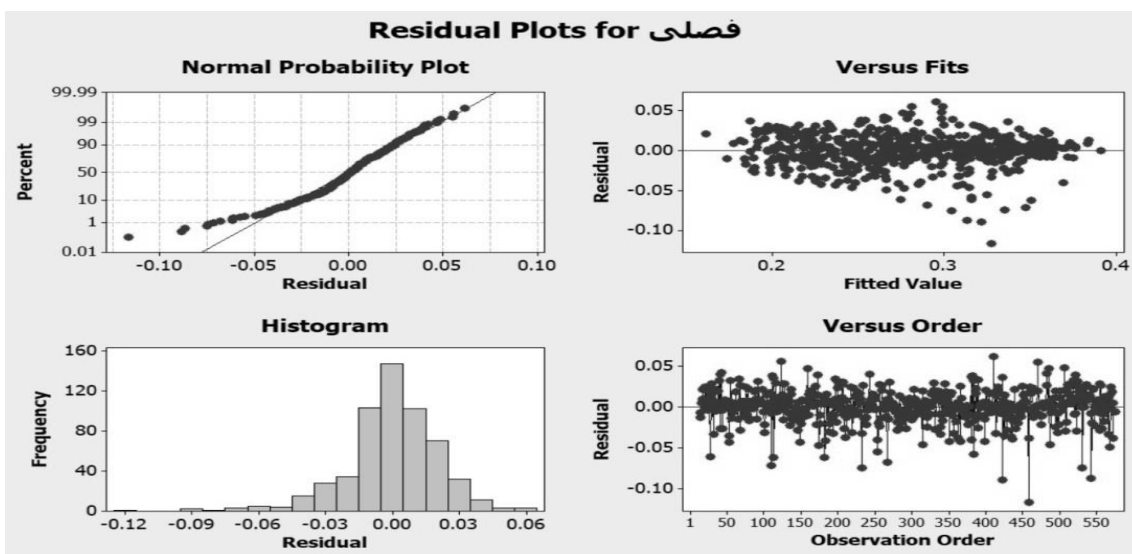
ابتدا برای هر مدل منتخب، آزمون پرت مانتو انجام و سپس نمودارهای مختلف مربوط به باقیمانده بررسی می‌شود. نتایج آزمون پرت مانتو برای مدل $SARIMA(4,0,4)(1,1,1)_{12}$ مطابق جدول (۶) است:

جدول (۶): نتایج آزمون پرت مانتو برای مدل $SARIMA(4,0,4)(1,1,1)_{12}$

Lag	۱۲	۲۴	۳۶	۴۸
Chi-Square	۶/۹	۱۹/۶	۲۹/۷	۴۶/۴
P-Value	۰/۰۳۱	۰/۱۴۴	۰/۲۷۹	۰/۱۶۳

می‌توان فرض صفر بودن همه خودهمبستگی‌ها را پذیرفت. همچنین گراف‌های حاصل از باقیمانده مدل مذکور در شکل (۷) نیز حکایت از مناسب بودن و کیفیت مدل دارد.

در مدل‌های فصلی، تأخیر یا گام زمانی را برای بررسی مدل اندکی بزرگ‌تر در نظر می‌گیرند. همان‌طور که ملاحظه می‌شود مقدار p-value برای تمامی تأخیرهای فوق حتی تا گام ۴۸ تقریباً بیشتر از ۰/۰۵ است؛ بنابراین



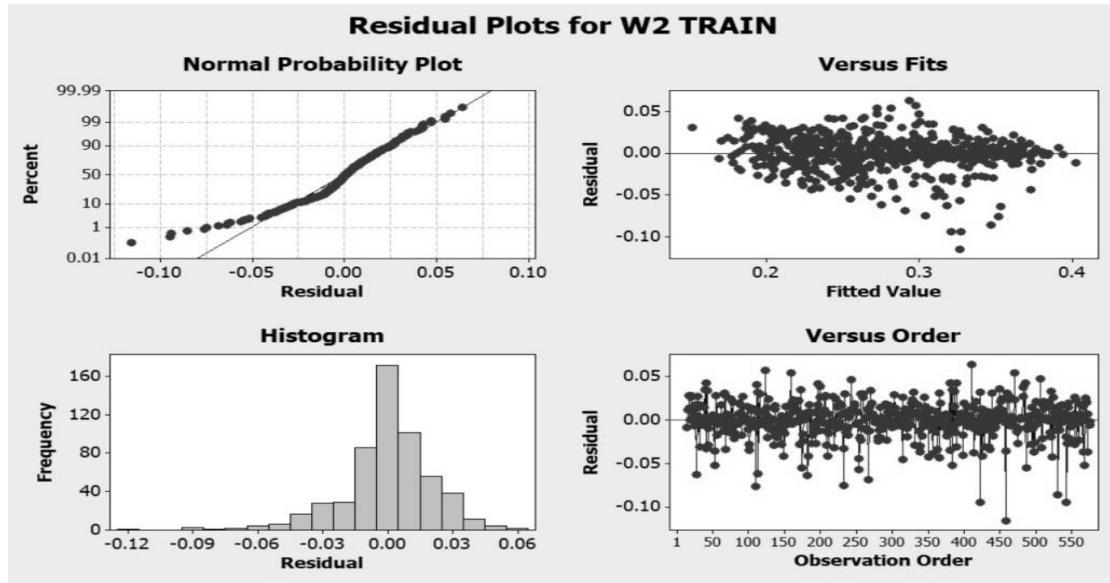
شکل (۷): بررسی نمودار باقیمانده‌های مدل $SARIMA(4,0,4)(1,1,1)_{12}$

جدول (۸): نتایج آزمون پرت مانتو برای مدل

$(1,1,1)(1,1,1)_{12}$

Lag	۱۲	۲۴	۳۶	۴۸
Chi-Square	۱۲/۲	۲۷	۳۷/۲	۵۴/۴
P-value	۰/۰۹۳	۰/۱۰۴	۰/۲۰۴	۰/۱۱۴

نتایج آزمون پرت مانتو برای مدل $SARIM(1,1,1)(1,1,1)_{12}$ در جدول (۷) و نمودارهای مختلف باقیمانده این مدل در شکل (۸) آورده شده است. نتایج بررسی‌های فوق نیز حکایت از مناسب بودن و کیفیت این مدل دارد.



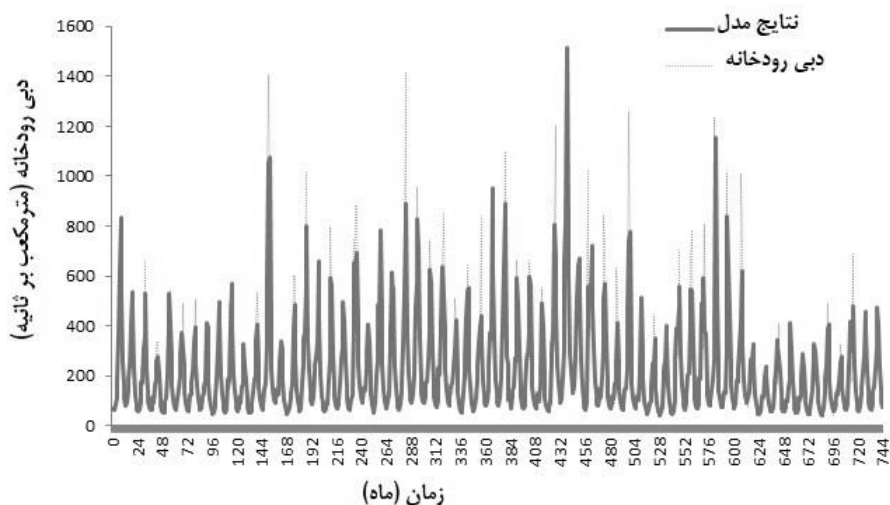
شکل (۸): بررسی نمودار باقیمانده‌های مدل $SARIMA(1,1,1)(1,1,1)_2$

به مدت ۲۴ ماه پیش‌بینی صورت گرفته که نتایج آن در شکل (۹) آورده شده است. همان‌گونه که در معیارهای ارزیابی مدل (جدول ۵) آورده شد نتایج صحت سنجی مدل، مطابقت خوبی با واقعیت داشته که این موضوع در شکل (۹) نیز قابل مشاهده است

پیش‌بینی

با انتخاب بهترین مدل مطرح شده یعنی مدل $SARIMA(4,0,4)(1,1,1)_{12}$ و بر اساس روابط (۱۹) - (۱۵)، الگوی دبی جریان رودخانه دز طبق رابطه (۲۱) به دست آمد. با برازش آن بر روی سری زمانی اصلی،

$$(1 - 0.03B^{12})(1 - 0.254B - 0.220B^2 - 0.1B^3 - 0.176B^4)(1 - B^{12})X(t) = (1 + B^{12})(1 - 0.347B - 0.029B^2 - 0.026B^3 + 0.146B^4)\varepsilon(t) \quad (21)$$



شکل (۹): نمودار سری زمانی اصلی و سری پیش‌بینی شده توسط مدل منتخب

نتیجه گیری

۳- بهترین مدل حاصل از سری زمانی دبی رودخانه با تفاضل گیری فصلی مدل $SARIMA(4,0,4)(1,1,1)_{12}$ و برای مدل منتخب سری زمانی با تفاضل گیری توأم مدل $SARIMA(1,1,1)(1,1,1)_{12}$ انتخاب گردید.

۴- برای مدل $SARIMA(4,0,4)(1,1,1)_{12}$ نتایج معیارهای ارزیابی $RMSE$ ، MAE و R در قسمت های واسنجی به ترتیب برابر $121/8$ ، $67/2$ و $0/85$ و در قسمت اعتبارسنجی به ترتیب برابر $154/4$ ، $92/47$ و $0/61$ قرار گرفت.

۵- برای مدل $SARIMA(1,1,1)(1,1,1)_{12}$ نتایج معیارهای ارزیابی $RMSE$ ، MAE و R در قسمت های واسنجی به ترتیب برابر $122/81$ ، $67/7$ و $0/85$ و در قسمت اعتبارسنجی به ترتیب برابر $170/17$ ، $107/7$ و $0/6$ قرار گرفت.

۶- نتایج نشان دادند برای هر دو مدل منتخب سری حاصل از تفاضل فصلی و تفاضل ترکیبی، دقت مدل در قسمت واسنجی بسیار بیشتر از اعتبارسنجی بوده و به طور کلی از میان هر دو مدل، مدل حاصل از تفاضل گیری فصلی دقت بهتری را نشان داد. نتایج پژوهش حاضر پتانسیل استفاده برای مناطق مشابه را دارا می باشد.

رودخانه دز یکی از مهمترین رودخانه ها در حوزه منابع آب کشور است به گونه ای که چندین ایستگاه هیدرومتری در طول رودخانه و به طور مستمر در حال پایش دبی آن هستند. در پژوهش حاضر به بررسی ایستایی اولیه سری زمانی دبی رودخانه دز پرداخته شد و تأثیر تفاضل گیری های مختلف غیر فصلی، فصلی و ترکیبی بر ایستایی سری زمانی و نتایج حاصل از مدل های منتخب آن به طور کامل بررسی گردید. خلاصه نتایج به شرح زیر است:

۱- سری حاصل از تفاضل گیری غیر فصلی همانند سری بدون تفاضل فاقد ایستایی لازم بود. نکته حائز اهمیت در این پژوهش، ناتوان بودن آزمونی قوی مانند دیکی- فولر تعمیم یافته در تشخیص نایستایی آن ها است. از بین دو سری دیگر یعنی سری با تفاضل توأم و سری با تفاضل فصلی، سری با تفاضل گیری توأم دارای بیشترین ایستایی بود.

۲- ایستایی بیشتر داده ها دستیابی به دقیق ترین مدل را آسان کرد به این ترتیب که به تعداد بسیار کمتری مدل برای کنترل حالات مختلف نیاز بود.

منابع

1. Dabral PP, Murry MZ. Modelling and forecasting of rainfall time series using SARIMA. Environmental Processes. 2017; 4(2):399-419.
2. Salmani MH, Jajaei ES. Forecasting models for flow and total dissolved solids in Karoun river-Iran. Journal of Hydrology. 2016; 535:148-59.
3. Haghizadeh A., Yousefi H., Yarahmadi Y., Nourmohammadi P., Alijani R. Prediction and analysis of river water quality parameters using ARIMA time series model In the basin of Kahman's river. Journal of Ecohydrology. 2017; 4(1):65-73 (IN PERSIAN).
4. Afrifa-Yamoah E. Application of ARIMA Models in Forecasting Monthly Average Surface Temperature of Brong Ahafo Region of Ghana. International Journal of Statistics and Applications. 2015; 5(5):237-46.
5. Yang JH, Cheng CH, Chan CP. A Time-Series Water Level Forecasting Model Based on Imputation and Variable Selection Method. methods. 2017; 29:32.



6. Bazrafshan, O., Chashmberah A., Holisaz A. Evaluation of time series models in forecasting pan evaporation in different climates of Hormozgan province. *Watershed Engineering and Management*. 2017; 9(3): 250-261.
7. Bazrafshan O, Salajegheh A, Bazrafshan J, Mahdavi M, Fatehi Maraj A. Hydrological Drought Forecasting using ARIMA Models (Case Study: Karkheh Basin). *Ecopersia*. 2015; 3(3):1099-117.
8. Patle GT, Singh DK, Sarangi A, Rai A, Khanna M, Sahoo RN. Time series analysis of groundwater levels and projection of future trend. *Journal of the Geological Society of India*. 2015; 85(2):232-42.
9. Al-Juboori AM, Guven A. A stepwise model to predict monthly streamflow. *Journal of Hydrology*. 2016; 543:283-92.
10. Elganiny MA, Eldwer AE. Comparison of Stochastic Models in Forecasting Monthly Streamflow in Rivers: A Case Study of River Nile and Its Tributaries. *Journal of Water Resource and Protection*. 2016;8(02):143.
11. Seyedian S.M., Soleiman M., Kashani M. River flow forecasting using data mining and time series. *Journal of Ecohydrology*. 2014;1(3):167-79 (IN PERSIAN)
12. Brockwell P.J., Davis, R.A. *Time series: theory and methods*. Second edition. NY: Springer Science & Business Media; 1991.
13. Moeeni H, Bonakdari H, Fatemi SE. Stochastic model stationarization by eliminating the periodic term and its effect on time series prediction. *Journal of hydrology*. 2017 1; 547:348-64.
14. Pena, D., Tiao, G.C., and Tsay, R.S. *A course in time series analysis*. NY: John Wiley & Sons; 2011.
15. Khademi M., Moeeni H., Bonakdari H., Ebtehaj I. The effect of differencing in stationery and accuracy of time series in predicting of lake level. *Journal of Water and Soil Conservation*. 2017;24(3):59-76(IN PERSIAN).
16. Salas J.D., Delleur J.W., Yevjevich V., Lane W.L. *Applied modeling of hydrologic time series*. Colorado: Water Resources Publication; 1980.
17. Bagherpoor M., Seyedian S. M., Fathabadi A. H., Mohamadi A. Study of Mann-Kendall Test Performance in Detecting the Series of Autocorrelation. *Watershed Engineering and Management*. 2017;11(36):11-21(IN PERSIAN).
18. Cryer J.D., Chan K.S. *Time Series Analysis with Applications in R*. Second Ed. NY: Springer; 2008.



Assessing the Improvement of Stability and Precision of Predicting the Time Series Models of River Flow under the influence of Differential Approach (Case study: Dez River)

Mohsen Moslemzadeh¹, Saeed Farzin^{2*}, Hojat Karami³

Abstract

The purpose of this study was to investigate the effect of seasonal, non-seasonal and combined differentiations on the time series stability of Dez River's monthly discharge for 60 years. Also, the effect of the stability level on the performance of SARIMA models in the prediction of the time series is investigated from different aspects. Out of 720 data, 80% are taken into account for the calibration and the rest for the validation period. First, the stability of the variance of time series was investigated and the necessary transformations were applied to stabilize the variance. Then, with the help of seasonal Mann-Kendall test, the average stability was evaluated and its results have been used to assess the existence of the process and the need for differentiation. Using seasonal, non-seasonal and combined differentiations, three new series have been created, whose stability with the original series have been investigated by analyzing the ACF graph and the generalized Dickey-Fuller test. In the following, the type and number of parameters required in the models are determined for each scenario. The results indicate that when using combined differentiation, the number of models needed for review is severely reduced so that for modeling of series with seasonal differentiation, 196 and for modeling of combined differentiation series, only 16 models need to be investigated. While the outcomes of the best model of both types of differentiated series were almost the same, the best model of the seasonal differencing series, with the criteria of the assessment of MAE= 92.4, RMSE = 154.4 and R = 0.61 showed its superiority over the other models. Finally, the selected model namely SARIMA (4, 0, 4) (1, 1, 1)₁₂ has been used to predict the river flow in the next 24 months.

Keywords: Differentiation, Stability, Prediction, River Flow, SARIMA

¹Ph. D student of Water Resources Engineering and Management, Faculty of Civil Engineering, Semnan University, Semnan, Iran. Email: moslemzadeh.m @semnan.ac.ir

²Assistant Professor, Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan, Iran. saeed.farzin@semnan.ac.ir

³Assistant Professor, Department of Water Engineering and Hydraulic Structures, Faculty of Civil Engineering, Semnan University, Semnan, Iran. Email: hkarami@semnan.ac.ir