

## تحلیل توانایی مدل میانگین متحرک جامع خود همبسته برای پیش‌بینی دو سال آینده جریان روزانه ورودی به مخزن سد دز

محمد ابراهیم بنی حبیب<sup>۱</sup>، ریحانه بندری<sup>۲</sup>، سید سعید موسوی ندوشنی<sup>۳</sup>

تاریخ دریافت: ۹۰/۰۵/۲۶

تاریخ پذیرش: ۹۰/۱۲/۲

### چکیده

یکی از روش‌های معمول در پیش‌بینی جریان رودخانه‌ها، مدل‌های سری زمانی می‌باشند. در این تحقیق به منظور پیش‌بینی آبدهی روزانه ایستگاه تله زنگ واقع در بالادست سد دز از مدل میانگین متحرک جامع خود همبسته (ARIMA) استفاده شده است. با توجه به اینکه این داده‌ها دارای نوسانات فصلی می‌باشند، با بهره‌گیری از سری فوریه، شاخص‌های آماری آن‌ها، نظیر میانگین و انحراف معیار برای دوره ۲۸ ساله با پیوند ۳۶۰ روزه برآورد شدند. سپس، داده‌های مشاهداتی آبدهی روزانه، توسط این شاخص‌های آماری استاندارد شدند. بررسی داده‌های استاندارد شده، نشان داد که روند فصلی داده‌ها، توسط عوامل محاسبه شده‌ی سری فوریه حذف شده است. در این تحقیق، برازش مدل‌های مختلف میانگین متحرک جامع خود همبسته به داده‌های استاندارد شده بررسی شده و در نهایت با استفاده از معیار آکائیک و در نظر گرفتن حداقل تعداد عوامل مدل‌ها، بهترین مدل انتخاب گردید. نتایج پیش‌بینی توسط مدل انتخابی نشان داد که این مدل توانسته است به طور نسبی روند آبدهی متوسط روزانه‌ی ورودی به مخزن سد دز برای دو سال آینده پیش‌بینی نموده و در مقایسه نتایج آن با تحقیقات گذشته نشان می‌دهد که میانگین قدر مطلق خطای نسبی پیش‌بینی آبدهی روزانه از ۳/۱۲٪ به ۰/۶٪ تنزل پیدا کرده و طول دوره پیش‌بینی از ده روز به دو سال افزایش یابد.

واژه‌های کلیدی: آریما (ARIMA)، پیش‌بینی جریان روزانه، روند فصلی داده‌ها، سری فوریه، سد دز.

<sup>۱</sup> نویسنده مسئول، دکترای عمران - مهندسی آب، استادیار گروه مهندسی منابع آب دانشگاه تهران، پردیس ابوریحان Banihabib@ut.ac.ir

<sup>۲</sup> دانشجوی کارشناسی ارشد رشته مهندسی منابع آب، دانشگاه تهران، پردیس ابوریحان R\_bandary@ut.ac.ir

<sup>۳</sup> استادیار گروه مهندسی آب، دانشگاه صنعت آب و برق (شهید عباسپور) s\_mousavi@pwut.ac.ir

## مقدمه

مدل‌های مختلفی به داده‌های ماهانه برازش داده شد و در نهایت با استفاده از شاخص‌های مقایسه‌ای نظیر  $MARE^2$  و  $CE^3$  بهترین مدل انتخاب شد (کاستالانو-مندز، ۲۰۰۴). بنی‌حبيب و ولی‌پور در سال ۱۳۸۷ ارزیابی مقایسه‌ای مدل‌های آرما، آرما و مدل خود همبسته شبکه عصبی مصنوعی را در پیش‌بینی جریان ماهانه ورودی به مخزن سد دز انجام دادند و در نهایت مدل شبکه عصبی مصنوعی دینامیک به عنوان بهترین مدل در پیش‌بینی جریان ماهانه ورودی به مخزن سد دز انتخاب شد. کمالی و همکاران در سال ۱۳۸۵ اقدام به پیش‌بینی جریان ماهانه ورودی به مخزن سد شهید عباسپور با بهره‌گیری از مدل SARIMA نمودند. توفیک<sup>۴</sup> در سال ۲۰۰۲ تقابل پیش‌بینی خطی و غیرخطی را در جریان ماهانه رودخانه نیل مورد بررسی قرار داد. از آنجا که مصر تقریباً به طور کلی به خاطر تأمین حدود ۹۵٪ از آب‌های مورد نیاز خود وابسته به رودخانه نیل است، سد بلند آسوان در بالادست رودخانه نقش مهمی دارد. به طوری که با هر تصمیمی برای رهاسازی، آب از دست رفته بازیافت نشده و مقدار قابل توجهی آب از دست خواهد رفت. بنابراین، پیش‌بینی‌های بلندمدت و کوتاه‌مدت جریان نیل برای مدیریت بهتر مخزن دارای اهمیت فراوانی است. مدل‌های خود همبسته متعددی برای پیش‌بینی جریان ورودی به مخزن سد آسوان استفاده شدند. ولی اکثر این مدل‌ها در پیش‌بینی شکست خوردند، چرا که نقطه پیک جریان در ماه‌های جولای، آگوست و سپتامبر به شدت تغییر می‌کرد. به عقیده او این عدم دقت در پیش‌بینی به خطی بودن مدل‌های خودهمبسته برمی‌گردد. به همین منظور از مدل‌های شبکه عصبی مصنوعی به جهت ساختار غیرخطی استفاده کرد (توفیک، ۲۰۰۲). نتایج نشان داد که غیرخطی بودن مدل‌های شبکه عصبی مصنوعی در بعضی موارد منجر به بهبود دقت پیش‌بینی آبدی شده است.

مروری بر تحقیقات پیشین نشان می‌دهد که پیش‌بینی آبدی ماهانه توسط سری‌های زمانی در مقایسه با شبکه عصبی مصنوعی از دقت پایین‌تری برخوردار است و در مقیاس روزانه نیز پیش‌بینی جریان با استفاده از سری‌های زمانی ترکیب شده با رگرسیون خطی انجام شده

استفاده مناسب از آب مخازن سدها به عنوان تأمین کننده اصلی آب بخش کشاورزی، شرب و صنعت در ایران در پرتو مدیریت و بهره‌برداری بهینه از این مخازن امکان‌پذیر است و این خود بدون اطلاع از میزان جریان ورودی به مخازن در دوره‌های زمانی آتی ممکن نمی‌باشد. در این راستا، برآورد هرچه دقیق‌تر مقدار جریان ورودی به مخازن، به دلیل نقشی که در مدیریت و بهره‌برداری از مخازن و تولید انرژی برقی دارد، از موضوعات با اهمیت در مهندسی منابع آب محسوب می‌شود. لذا به منظور مدیریت بهینه منابع آب، تأمین بهینه نیازهای آبی، بهره‌برداری و مدیریت بهینه سیلاب توسط مخزن سد، استفاده از پیش‌بینی آبدی روزانه‌ی رودخانه‌ها، امری ضروری می‌باشد (دونگ و همکاران، ۲۰۰۶ و پارکر و همکاران، ۲۰۰۵).

پیش‌بینی‌های موجود از جریان ورودی به مخازن با دو شاخص قابل بررسی است. شاخص اول مقیاس زمانی داده‌های آبدی نظیر آبدی متوسط روزانه، ماهانه و سالانه می‌باشد. شاخص دوم زمان پیش‌بینی است که به مدت زمانی گفته می‌شود که از قبل بتوان آبدی را پیش‌بینی نمود. بررسی تحقیقات پیشین نشان می‌دهد که اکثر کارهایی که تاکنون در زمینه پیش‌بینی جریان توسط مدل‌های سری زمانی انجام شده در مقیاس ماهانه بوده و به داده‌های روزانه کمتر توجه شده است (شالامو، ۲۰۰۹، بنی‌حبيب، ولی‌پور و بهبهانی، ۱۳۸۷، وانگ و همکاران، ۲۰۰۵ و نواکز و همکاران، ۱۹۸۵). معدودی پیش‌بینی آبدی در مقیاس روزانه انجام شده که حدکثر زمان پیش‌بینی ده روز بوده است (وانگ، ۲۰۰۶). همچنین کمترین میانگین قدر مطلق خطای نسبی گزارش شده پیش‌بینی آبدی روزانه ۳/۱۲ بوده است (کاستالانو-مندز، ۲۰۰۴).

تحقیقات گسترده‌ای در دنیا در زمینه پیش‌بینی جریان انجام شده است که این موضوع نشان‌گر اهمیت موضوع پیش‌بینی جریان ورودی به سدها می‌باشد. کاستالانو-مندز<sup>۱</sup> و همکاران در سال ۲۰۰۴ اقدام به مدل‌سازی بارش-رواناب در مقیاس ماهانه و روزانه با استفاده از مدل‌های سری زمانی نمودند. بدین منظور

<sup>2</sup> Mean absolute relative error<sup>3</sup> Coefficient of efficiency<sup>4</sup> Tawfik<sup>1</sup> Castellano-Mendez



نتایج حاصل از بررسی روابط همبستگی جهت بازسازی داده‌ها، در جدول (۱) ارائه شده است. با توجه به حداقل بودن MARE مدل ایستگاه تنگ پنج بختیاری و مناسب بودن ضریب تعیین این مدل، مدل یک متغیره تنگ پنج بختیاری برای بازسازی انتخاب شد. اگرچه ضریب تعیین در مدل دو متغیره نسبت به مدل یک متغیره با ایستگاه تنگ پنج بالاتر است، اما بزرگتر بودن میانگین قدر مطلق خطای نسبی (MARE) در مدل دو متغیره، عامل انتخاب مدل یک متغیره با استفاده از ایستگاه تنگ پنج بختیاری برای بازسازی داده‌ها می‌باشد.

شده در روز  $t$ ام،  $Q_{oi}$  دبی مشاهداتی در روز  $t$ ام،  $n$  تعداد داده‌ها،  $\bar{Q}_{obs}$  متوسط دبی مشاهداتی هستند.

$$MARE = \left( \sum \frac{|Q_{oi} - Q_{ci}|}{Q_{oi}} \right) / n \quad (1)$$

$$R^2 = 1 - \frac{\sum_{i=1}^n (Q_{ci} - Q_{oi})^2}{\sum_{i=1}^n (Q_{oi} - \bar{Q}_{obs})^2} \quad (2)$$

جدول (۱): بررسی شاخص‌های آماری برای انتخاب ایستگاه مبنای بازسازی داده‌های ایستگاه تله زنگ

نوع مدل	مدل‌های یک متغیره	مدل دو متغیره
ایستگاه مبنای بازسازی	تنگ پنج سزار	رگرسیون ۲ متغیره با دو ایستگاه تنگ پنج بختیاری و تنگ پنج سزار
MARE	۰/۳۳	۰/۲۸
R <sup>2</sup>	۰/۸۳	۰/۹۰

حاضر به منظور نرمال نمودن داده‌ها از روش تبدیل توانی باکس-کاکس (رابطه ۳) استفاده شده است:

$$\begin{cases} Z_t = \frac{Z_t^\lambda - 1}{\lambda} & \lambda \neq 0 \\ Z_t^{(\lambda)} = \ln(Z_t) & \lambda = 0 \end{cases} \quad (3)$$

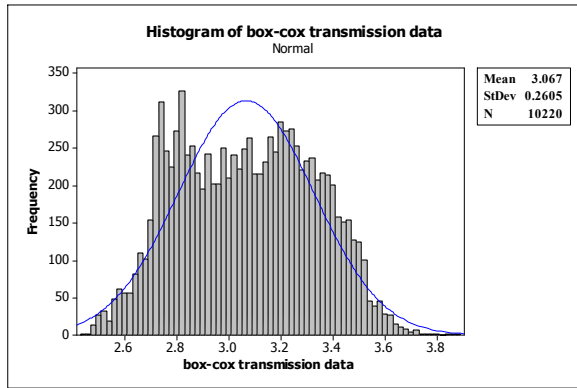
که در آن  $Z_t$  مقدار اولیه داده‌ها در زمان  $t$ ،  $\lambda$  پارامتر تبدیل و  $Z_t^{(\lambda)}$  مقدار تبدیل یافته نظیر داده‌ها است. پارامتر تبدیل  $\lambda$  بر اساس نمودار شکل (۳) در نقطه‌ای که log-likelihood ماکزیمم مقدار را داراست انتخاب می‌شود. در این شرایط  $\lambda = -0/22$  می‌باشد. اشکال (۴) و (۵) به ترتیب هیستوگرام داده‌ها قبل و بعد از تبدیل باکس کاکس را نشان می‌دهند. همانگونه که شکل (۴) نشان می‌دهد داده‌ها از توزیع نرمال پیروی نمی‌نمایند. اما با توجه به شکل (۵) داده‌ها بعد از تبدیل باکس-کاکس به فرم زنگوله‌ای در آمده و تا حد قابل قبول داده‌های روزانه نرمال شده‌اند.

### مراحل انجام پیش‌بینی آینده روزانه

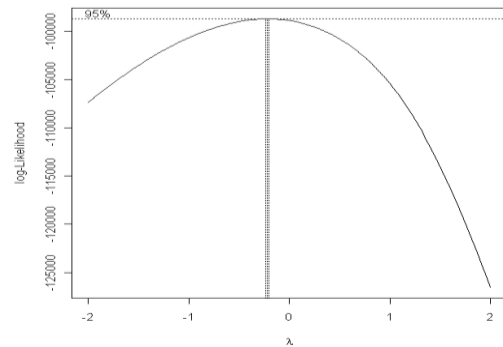
در این تحقیق به منظور پیش‌بینی آینده روزانه‌ی ورودی به مخزن سد دز در محل ایستگاه تله‌زنگ، طول دوره آماری آینده‌ی روزانه این ایستگاه از ابتدای سال ۱۳۵۷ تا پایان سال ۱۳۸۶ (دوره ۳۰ ساله) استفاده شده است. در واقع آمار مورد استفاده، شامل ۱۰۹۵۰ داده روزانه می‌باشد که از این تعداد ۱۰۲۲۰ داده اولیه (ابتدای فروردین ۱۳۵۷ تا پایان اسفند ۱۳۸۴) به منظور تعیین عوامل مدل ARIMA استفاده شده‌اند. ۷۳۰ داده پایانی (ابتدای فروردین ۱۳۸۵ تا پایان اسفند ۱۳۸۶) برای صحت‌سنجی پیش‌بینی مورد استفاده قرار گرفتند.

به منظور تجزیه و تحلیل سری زمانی، بررسی نرمال بودن داده‌ها امری ضروری می‌باشد. نرمال بودن داده‌ها به این دلیل اهمیت دارد که تئوری سری‌های زمانی بر اساس نرمال بودن داده‌ها توسعه یافته و در صورتی که داده‌ها نرمال نباشند بایستی از روش‌های مختلف چون لگاریتم‌گیری، جذر، نمایی و یا تبدیل توانی باکس-کاکس<sup>۱</sup> آن‌ها را نرمال نمود (ونابلس، ۲۰۱۱). در مقاله

<sup>۱</sup> Box-Cox

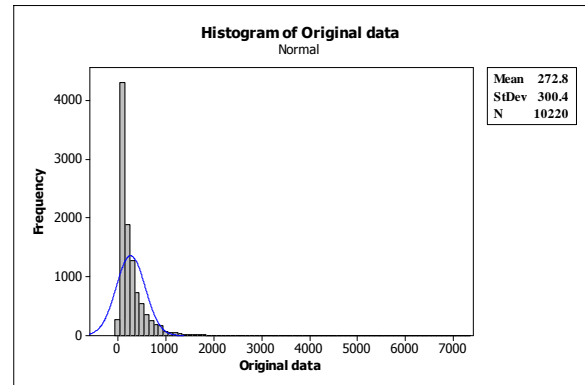


شکل (۵) هیستوگرام داده‌ها بعد از تبدیل باکس-کاکس



شکل (۳): نمودار باکس-کاکس

پس از نرمال نمودن داده‌ها، به منظور استاندارد سازی داده‌ها، شاخص‌های آماری (میانگین و انحراف معیار داده‌های مشاهداتی) توسط سری فوریه محاسبه شدند. سری فوریه با رابطه (۴) بیان می‌شود. فرض کنید که  $u_\tau$  برآورد نمونه‌ای از عوامل تناوبی جامعه  $u_\tau$  باشد. در این فرمول  $\bar{u}$  میانگین  $u_\tau$  است.  $A_j$  و  $B_j$  ضرایب سری فوریه هستند که از روابط (۵) تا (۷) محاسبه می‌شوند.  $J$  هارمونی و  $h$  مجموع هارمونی است که برابر  $w/2$  و یا  $(w-1)/2$  بر حسب زوج و فرد بودن  $w$  می‌باشد.  $w$  متغیری است که به نوع مقیاس زمانی داده‌ها وابسته است. برای مثال در مقیاس داده‌های روزانه  $w=365$  و  $h=182$  است (سالاس، ۱۹۸۰).



شکل (۴): هیستوگرام داده‌ها قبل از تبدیل باکس-کاکس

$$\widehat{u}_T = \bar{u} + \sum_{j=1}^h \left[ A_j \cos\left(\frac{2\pi j \tau}{w}\right) + B_j \sin\left(\frac{2\pi j \tau}{w}\right) \right] \quad (4)$$

$$\tau = 1, \dots, w$$

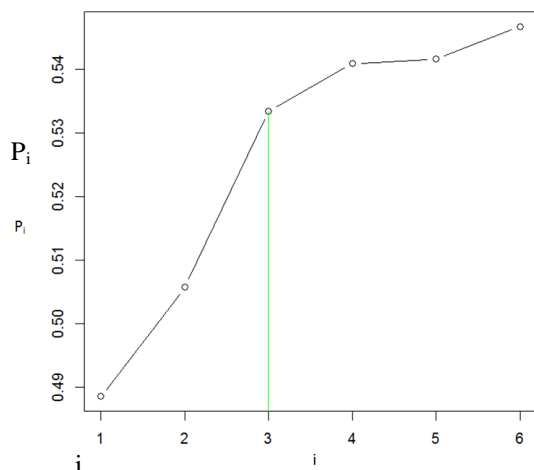
$$j = 1, \dots, h, B_j = \frac{2}{w} \sum_{\tau=1}^w u_\tau \sin\left(\frac{2\pi j \tau}{w}\right) \quad (5)$$

$$j = 1, \dots, h, A_j = \frac{2}{w} \sum_{\tau=1}^w u_\tau \cos\left(\frac{2\pi j \tau}{w}\right) \quad (6)$$

$$\bar{u} = \frac{1}{w} \sum_{\tau=1}^w u_\tau \quad (7)$$

بدست می‌آید.  $p_i$  که عددی بدون بعد است را بر حسب  $i$  رسم می‌نماییم. شکل حاصل را تناوب نگار تجمعی می‌نامند. شکل‌های (۶) و (۷) تناوب نگار تجمعی میانگین و انحراف معیار داده‌ها هستند. ضابطه گرافیکی برای

برای انتخاب هارمونی‌ها از تناوب نگار تجمعی استفاده می‌کنیم. برای این کار مربع انحراف از میانگین از رابطه (۸) محاسبه می‌شود. آنگاه مقدار  $MSD(j)$  از رابطه (۹) محاسبه می‌شود. پس از محاسبه  $MSD(j)$  نسبت  $p_i$  از رابطه (۱۰)



شکل (۷) تناوب نگار تجمعی انحراف معیار داده‌ها

پس از استاندارد نمودن داده‌ها بایستی ایستایی داده‌ها مورد بررسی قرار گیرد. تعریف ایستایی را به طور موجز می‌توان ثابت بودن میانگین، واریانس و دیگر عوامل آماری تعریف کرد. برای تشخیص ایستایی یا غیرایستایی یک سری از تابع خود همبستگی<sup>۱</sup> (ACF) استفاده می‌شود. از آنجا که ضریب همبستگی (r) تابعی از تأخیرها (k) است به آن تابع خود همبستگی گویند و آن را با ACF نشان می‌دهند. اگر همبستگی بین یک سری را با سری‌های تشکیل شده از گام‌های مختلف زمانی آن بدست آوریم ( $r_k$ ) و ضرایب این همبستگی را در مقابل گام‌های تأخیر زمانی ترسیم کنیم، نمودار حاصل را خود همبستگی نگار می‌نامند. رابطه (۱۲) تابع خود همبستگی با تأخیر k را نشان می‌دهد (سالاس، ۱۹۸۰).

(۱۲)

$$r_k = \frac{\sum_{i=1}^{n-k} (x_i - \bar{x})(x_{i+k} - \bar{x})}{\sum_{i=1}^n (x_i - \bar{x})^2} \quad -1 \leq r_k \leq 1$$

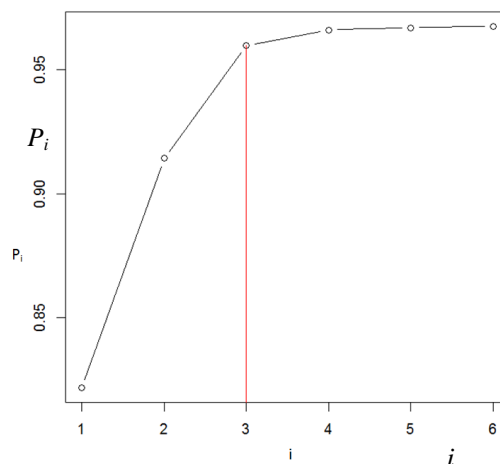
که در این رابطه  $r_k$  مقدار تابع خود همبستگی سری زمانی با تأخیر  $k$  و  $x_{i+k}$  و  $x_i$  ترتیب مقادیر متغیرها یا داده‌های سری زمانی در مرحله زمانی  $i$  و مرحله با تأخیر زمانی  $k$ ، مقدار میانگین مربوط به متغیرها باشد. با توجه به اینکه محدوده مجاز ACF،  $2/\sqrt{n}$  می‌باشد که در آن  $N$  تعداد داده‌ها است (سالاس، ۱۹۸۰). اگر حدود ACF پس از نزول در این محدوده قرار گیرد آن

انتخاب هارمونی‌های معنی‌دار با استفاده از نمودار تناوب نگار تجمعی به شرح زیر است. به منظور انتخاب تعداد هارمونی معنی‌دار در محل تقاطع دو بخش با شیب ملایم و شیب تند، خطی عمود بر محور افقی رسم می‌نمائیم، عددی که محور افقی نشان می‌دهد تعداد هارمونی معنی‌دار می‌باشد. در مقاله حاضر تعداد هارمونی معنی‌دار با توجه به شکل‌های (۶) و (۷) و چگونگی افزایش داده‌ها، برای میانگین داده‌ها عدد ۳ و برای انحراف معیار داده‌ها عدد ۴ بدست آمد که تقاطع آن‌ها با محور افقی بر روی شکل‌های (۶) و (۷) نشان داده شده‌اند

$$MSD(u) = \frac{1}{w} \sum_{\tau=1}^w u_{\tau} - \bar{u} \quad (۸)$$

$$MSD(j) = \frac{1}{2} (A_j^2 + B_j^2) \quad j = 1, \dots, h \quad (۹)$$

$$P_i = \frac{\sum_{j=1}^i MSD(j)}{MSD(u)} \quad i = 1, \dots, h \quad (۱۰)$$



شکل (۶) تناوب نگار تجمعی میانگین داده‌ها

استاندارد نمودن داده‌ها با رابطه (۱۱) صورت گرفته است. پس از محاسبه تعداد هارمونی‌های معنی‌دار با استفاده از سری فوریه، میانگین و انحراف معیار را محاسبه نموده و میانگین بدست آمده از سری فوریه ( $\mu$ ) را از داده‌های اولیه ( $Xx$ ) کسر نموده و بر انحراف معیار  $\sigma$  تقسیم می‌نماییم. با استفاده از این روش داده‌ها استاندارد می‌شوند.

$$z = \frac{x - \mu}{\sigma} \quad (۱۱)$$

<sup>۱</sup> Auto Correlation Function

داده‌ها را ایستا در نظر می‌گیریم. با توجه به شکل (۸) مقادیر ACF سری زمانی داده‌های استاندارد شده در تأخیر ۴۰ به صفر نزدیک شده و در محدوده مجاز ACF واقع شده است. این امر نشان دهنده ایستا بودن داده‌ها می‌باشد. پس از استاندارد سازی داده‌ها و بررسی ایستایی داده‌ها، مرحله اول انتخاب مدل برتر، مرحله شناسایی آزمایشی می‌باشد که با استفاده از توابع خودهمبستگی (ACF) و خودهمبستگی جزئی (PACF) انجام می‌شود. خودهمبستگی جزئی (PACF) در واقع روشی دیگر برای عنوان نمودن وابستگی زمانی در ساختار یک سری زمانی می‌باشد. اگر  $\theta_k$  تابع خودهمبستگی جزئی سری زمانی با تأخیر  $k$  باشد، رابطه تابع خودهمبستگی جزئی (PACF) به صورت رابطه (۱۳) نشان داده می‌شود (سالاس، ۱۹۸۰):

$$\theta_k = \frac{r_k - \sum_{i=1}^{k-1} \theta_i(k-1)r_{k-i}}{1 - \sum_{i=1}^{k-1} \theta_i(k-1)r_i} \quad (13)$$

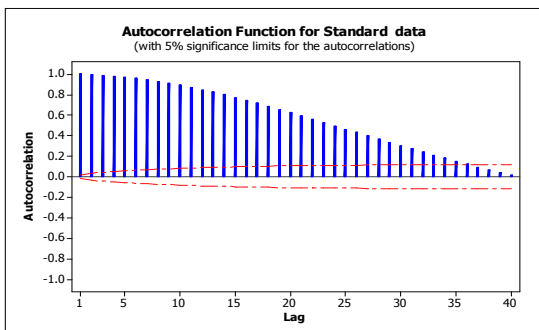
که در این رابطه  $\theta_k$ : مقدار تابع خود همبستگی جزئی سری زمانی با تأخیر  $k$  می‌باشد. شکل (۹) نمودار PACF داده‌ها را نشان می‌دهد. با توجه به اینکه نمودار ACF و PACF هر دو کاهش را نشان می‌دهند لذا داده‌ها ایستا بوده و نیازی به عمل تفاضل نیست. بنابراین معلوم می‌گردد که مدل مناسب مدل ARIMA (p,0,q) می‌باشد. در این حالت مرتبه جزء خودهمبسته و جزء میانگین متحرک که به ترتیب برابر با  $p$  و  $q$  می‌باشند روشن نیست، اما کاملاً منطقی است که بایستی از ARIMA (1,0,1) شروع نمود و چنانچه برازش مناسب نبود از مرتبه‌های دیگر استفاده نمود. فرمول مدل ARIMA (p,0,q) در رابطه (۱۴) ارائه شده است (کرایر و چان، ۲۰۰۸):

$$y_t = \theta_1 y_{t-1} + \dots + \theta_p y_{t-p} + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q} \quad (14)$$

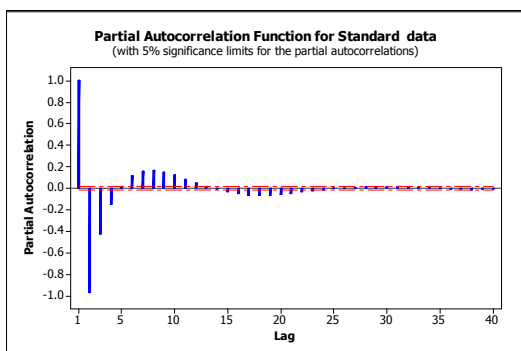
به طوری که در آن  $y_t$  و  $\varepsilon_t$  به ترتیب مقادیر واقعی و باقیمانده‌های مدل<sup>۲</sup> (نوفه سفید) در دوره  $t$

$$AIC = -2 \times \log\text{-likelihood} + 2 \times \text{Number of parameters} \quad (15)$$

در جدول (۲) مشخصات دو مدل برتر ارائه شده است. مدل‌های ARIMA (۵, ۰, ۴) و ARIMA (۵, ۰, ۵) کمترین مقدار AIC در بین مدل‌های برازش داده شده می‌باشند و با توجه به نزدیک بودن مقدار AIC آن‌ها و کم بودن تعداد عوامل مدل ARIMA (۵, ۰, ۴) نسبت به مدل ARIMA (۵, ۰, ۵)، مدل ARIMA (۵, ۰, ۴) به عنوان مدل برتر انتخاب شد.



شکل (۸): ACF سری زمانی داده‌های استاندارد شده



شکل (۹): PACF سری زمانی داده‌های استاندارد شده

<sup>3</sup> Parsimony

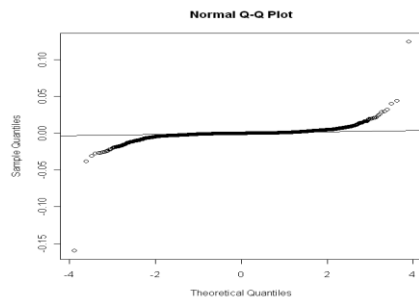
<sup>4</sup> Akaike Information Criterion

<sup>1</sup> Partial Auto Correlation Function

<sup>2</sup> White noise

برابر با  $i$  امین ACF باقیمانده‌ها است و  $p$  و  $q$  رتبه‌های مدل و  $n$  تعداد نمونه‌ها هستند. فرض صفر ( $H_0$ ) این است که ACF تمام باقیمانده‌ها برابر صفر و داده‌ها تصادفی هستند. P-value در نمودار لجونگ و باکس احتمال آن است که  $Q_K^*$  بزرگتر از  $Q$  مشاهداتی باشد. در این شرایط اگر p-value کوچکتر یا مساوی سطح معنی‌دار بودن آزمون یعنی ۵٪ باشد، فرض صفر ( $H_0$ ) رد می‌شود.

نتایج آماره لجونگ و باکس بررسی شده برای مدل  $(4, 0, 5)$  ARIMA در شکل (۱۱) ارائه شده است. با توجه به نمودار ACF باقیمانده‌ها و اینکه حدود ACF در محدوده مجاز قرار گرفته است، لذا باقیمانده‌ها ایستا هستند. همچنین باقیمانده‌ها هیچگونه روندی را نشان نمی‌دهند. همچنین با توجه به نمودار p-value در شکل (۱۱) و اینکه در مدل  $(4, 0, 5)$  ARIMA، p-value بزرگتر از سطح معنی‌دار بودن آزمون یعنی ۵٪ می‌باشد، فرض صفر قبول می‌شود و لذا باقیمانده‌ها تصادفی هستند. بدین ترتیب مدل  $(4, 0, 5)$  ARIMA مدل مناسبی است و می‌توان پیش‌بینی جریان را با استفاده از این مدل انجام داد.



شکل (۱۰): نمودار شاخص نرمال باقیمانده‌های مدل منتخب

### بحث روی نتایج پیش‌بینی

پس از انتخاب مناسب‌ترین مدل با استفاده از داده‌های مشاهداتی ۲۸ سال، پیش‌بینی جریان با استفاده از مدل  $(4, 0, 5)$  ARIMA برای روزهای مختلف در طی سال‌های ۱۳۸۵ و ۱۳۸۶ انجام شد. شکل (۱۲) نمودار آبدی‌های مشاهداتی، آبدی پیش‌بینی شده را نشان می‌دهد. همانطور که شکل (۱۲) نشان می‌دهد، مدل منتخب، توانسته است روند آبدی را به صورت قابل قبولی پیش‌بینی نماید. اما در نقاطی که آبدی‌ها صعود و نزول پی در پی دارند و رفتاری غیر خطی از خود نشان

### جدول (۲): ضرایب AIC و

پارامترهای مدل‌های برازش داده شده

مدل‌های برازش داده شده		پارامترها
ARIMA(5,0,4)	ARIMA(5,0,5)	
-۸۷۷۲۵/۹۷	-۸۷۷۳۴/۳۳	AIC Coefficient
-۳/۳۹	۳۱۸۰۹	$\theta_1$
-۳/۸۳	-۱/۸۲	$\theta_2$
۴۰۶۶۴	-۱/۸۱	$\theta_3$
۰/۷۷	۲۲۳۱۳	$\theta_4$
-۰/۴	-۰/۸۴	$\theta_5$
-۲/۰۱	-۱/۴۹	$\theta_1$
۴۰۵۴۴	-۰/۱۶	$\theta_2$
۰/۱۸	۱۳۱۵۰	$\theta_3$
-۰/۲۲	-۰/۶۹	$\theta_4$
-	۰/۰۷	$\theta_5$

پس از انتخاب مدل برتر در مرحله اولیه، بایستی آزمون نکویی برازش بر روی باقیمانده‌های مدل انجام شود تا از انتخاب مدل اطمینان حاصل شود. اولین موردی که به منظور بررسی باقیمانده‌ها انجام می‌شود، بررسی نرمال بودن باقیمانده‌هاست. شکل (۱۰) نمودار شاخص نرمال باقیمانده‌های مدل  $(4, 0, 5)$  ARIMA را نشان می‌دهد. از روی این نمودار می‌توان برازش مناسب داده‌ها بر خط شاخص نرمال و منطبق بودن اکثر داده‌ها بر خط شاخص نرمال را تشخیص داد. لازم به ذکر است که در دو سر خط شاخص نرمال، تعدادی از داده‌ها از خط دور می‌شوند. با توجه به اینکه قسمت عمده داده‌ها بر خط منطبق هستند و همچنین داده‌ها در مقیاس روزانه هستند و حاوی داده‌های سیلابی و کم آبی فراوانی هستند، این موضوع غیر قابل اجتناب به نظر می‌رسد. برای تکمیل آزمون‌ها می‌توان در مورد همبستگی پیاپی باقیمانده‌ها نیز بحث نمود. آماره لجونگ و باکس<sup>۱</sup> که این امر را به عهده دارد از رابطه (۱۶) بدست می‌آید (کرایر و چان، ۲۰۰۸).

$$Q_K^* = n(n+2) \sum_{i=1}^k \frac{n^2}{n-i} \quad (16)$$

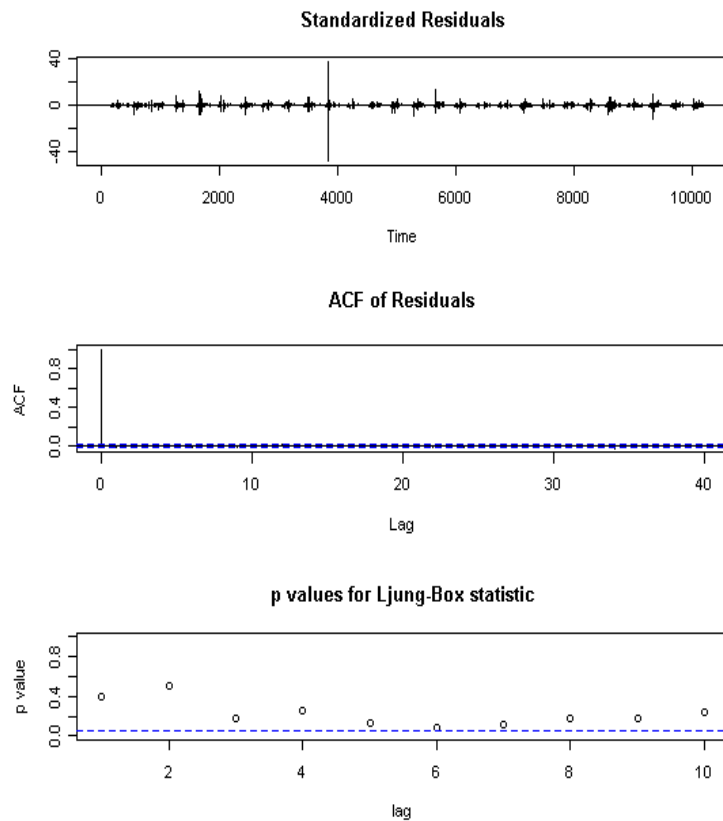
اگر این آزمون مثلاً برای مدل  $(p, 0, q)$  ARIMA به کار رود، آنگاه آماره  $Q_K^*$  دارای توزیع کای اسکوتر ( $x_{k-p-q}^2$ ) است. که در آن  $k$  تعداد تأخیرها و  $i$

<sup>۱</sup> Ljung-Box

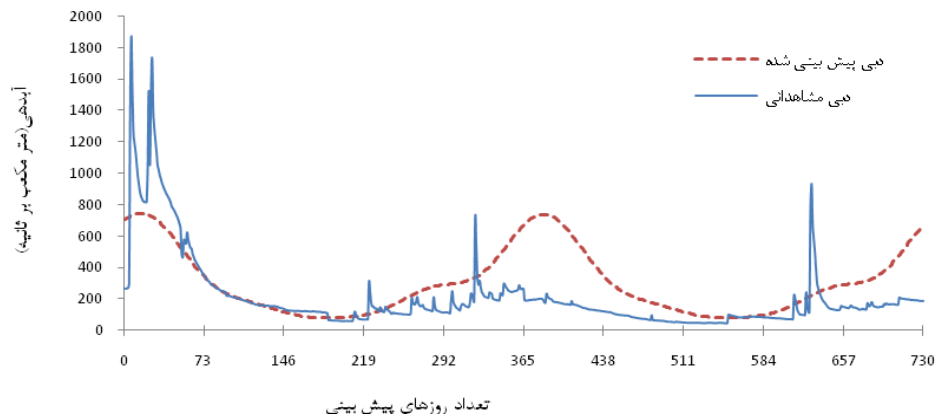


مقدار را دارا می‌باشد و به تدریج با پیش رفتن به سوی فصل گرما میزان خطای پیش بینی کاسته می‌شود و در فصل تابستان به کم‌ترین مقدار خود می‌رسد. پس از اتمام فصل تابستان و پیش رفتن به سوی فصل زمستان میزان خطاها افزوده می‌شود. با توجه به نمودار آبدهی پیش‌بینی شده مشاهده می‌گردد که در فصل بهار سال ۱۳۸۶ میزان خطا افزوده شده و روند تغییرات خطای پیش‌بینی مطابق سال ۱۳۸۵ تکرار می‌شود.

می‌دهند، پیش‌بینی مناسبی انجام نشده است. با توجه به نمودار آبدهی مشاهداتی و آبدهی پیش‌بینی شده که در شکل (۱۲) ارائه شده است. به منظور بررسی دقیق‌تر خطا در دوره پیش‌بینی شده نمودار میانگین قدر مطلق خطای نسبی آبدهی‌های روزانه پیش‌بینی شده توسط مدل  $ARIMA(5, 0, 4)$  در ماه‌های مختلف سال ۱۳۸۵ و ۱۳۸۶ در شکل (۱۳) ارائه شده است. مشاهده می‌شود که میزان خطای پیش‌بینی در فصل بهار سال ۱۳۸۵ ماکزیمم



شکل (۱۱): نمودار آماره لجونگ-باکس مدل  $ARIMA(5, 0, 4)$

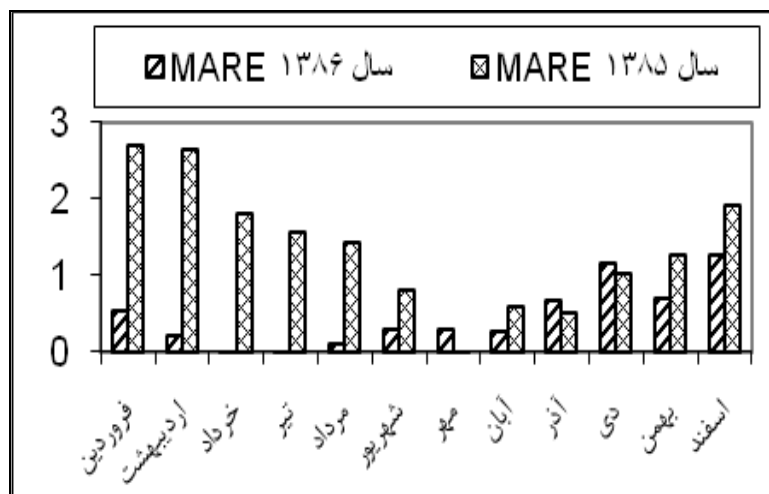


شکل (۱۲): نمودار آبدهی مشاهداتی و آبدهی پیش‌بینی شده توسط مدل پیشنهادی در سال‌های ۱۳۸۵ و ۱۳۸۶

پیش‌بینی تکراری است، اما میزان خطای پیش‌بینی با افزایش تعداد سال‌های پیش‌بینی بیشتر می‌شود و عملاً پیش‌بینی‌های طولانی مدت بیشتر از یک سال توصیه نمی‌شود.

میانگین قدرمطلق خطای نسبی در طی سال‌های ۱۳۸۵ و ۱۳۸۶، برابر با ۰/۶ می‌باشد. در مقایسه با تحقیقات صورت گرفته توسط کاستالونو-مندز (۲۰۰۴)، میانگین قدر مطلق خطای نسبی از ۳/۱۲ به ۰/۶ کاهش یافته است. کاهش چشمگیر در میزان خطا و افزایش زمان پیش‌بینی از ۱۰ روز به ۷۳۰ روز نشان دهنده کارایی سری فوریه در برآورد عوامل و مدل آریمای برازش داده شده نسبت به مدل (1) ARX می‌باشد.

با توجه به شکل (۱۳) سه ماه فصل تابستان کمترین خطای پیش‌بینی و سه ماه فصل زمستان بیشترین خطای پیش‌بینی را دارا می‌باشند. همچنین لازم به ذکر است که در حوضه دز عمده بارش‌ها در آخر فصل زمستان و شروع فصل بهار و پس از آن در فصل پاییز رخ می‌دهد. لذا شاید بتوان علت خطای بزرگتر پیش‌بینی را در زمستان، و به وقوع بارش‌ها نسبت داد. با افزایش احتمال بارندگی، میزان تغییرات دبی رودخانه شدیدتر بوده و وجود این تغییرات پی در پی، پیش‌بینی آبدهی را مشکل‌تر می‌نماید. اما در فصل تابستان به دلیل تغییرات کم آبدهی رودخانه پیش‌بینی با خطای کم‌تری انجام می‌شود. لذا با توجه به کلیه شرایط می‌توان اظهار نمود که اگرچه روند تغییرات خطای



شکل (۱۳): نمودار میانگین قدر مطلق خطای نسبی آبدهی‌های روزانه پیشنهادی شده توسط مدل پیشنهادی

نتایج پیش‌بینی نشان می‌دهد که با بکارگیری سری فوریه در برآورد عوامل و مدل (۵، ۰، ۴) ARIMA خطای پیش‌بینی از  $MARE = 3/12$  به  $MARE = 0/6$  تنزل پیدا کرده است. زمان پیش‌بینی نیز نسبت به تحقیقات پیشین از ده روز به دو سال افزایش پیدا کرده است. شایان ذکر است که اگرچه الگوی پیش‌بینی توسط مدل منتخب، الگوی مناسبی است و به خوبی توانسته است روند خطی آبدهی را در دوره ۷۳۰ روزه پیش‌بینی نماید، اما در نقاط پیک آبدهی، با توجه به غیر خطی بودن روند آبدهی‌ها پیش‌بینی مناسبی انجام نشده است. از سوی دیگر با توجه به اینکه پیش‌بینی داده‌ها در مقیاس روزانه انجام شده

## نتیجه گیری

با توجه به نمودار آبدهی مشاهداتی و آبدهی پیش‌بینی شده و برازش مناسب داده‌ها و همچنین مناسب بودن خطای پیش‌بینی، الگوی خطی پیش‌بینی توسط مدل (۵، ۰، ۴) ARIMA الگوی مناسبی است و توانسته است روند آبدهی را به صورت خطی پیش‌بینی نماید. اما با افزایش تعداد سال‌های پیش‌بینی میزان خطای پیش‌بینی چندین برابر شده است، افزایش چشمگیر خطای پیش‌بینی با افزایش تعداد سال‌های پیش‌بینی نشان می‌دهد که مدل منتخب برای پیش‌بینی‌های زیر یک سال مناسب است. در مقایسه با تحقیقات گذشته،

### تقدیر و تشکر

بدین وسیله از سازمان آب و برق خوزستان که طی قرارداد پژوهشی شماره ۱۴۳/۱۵۸۴۸۴ از این طرح تحقیقاتی حمایت کرده‌اند، تقدیر و تشکر بعمل می‌آید.

است و در داده‌های روزانه به دلیل اینکه آبدهی‌ها مانند داده‌های ماهانه از تقریباً روند یکنواختی پیروی نمی‌نمایند، عدم پیش‌بینی مناسب در نقاط حدی امری دور از ذهن نیست. همچنین همانگونه که پیش‌تر نیز اشاره شد، خطای پیش‌بینی در فصولی که بارندگی زیادی دارند نسبت به سایر فصول بیش‌تر است.

### منابع

۱. بنی‌حبیب، م. ا. و م. ولی‌پور. ۱۳۸۷. ارزیابی مقایسه‌ای مدل‌های آرما، آریما و مدل خودهمبسته شبکه عصبی مصنوعی در پیش‌بینی جریان ورودی به مخزن سد دز. اولین کنفرانس بین‌المللی بحران آب. اسفند ماه. دانشگاه زابل. ۸ صفحه.
2. Castellano-Mendez, M., W. Iez-Manteiga, M. Febrero-Bande, J. Prada- Sanchez, and R. Lozano Caldero. 2004. Modeling of the monthly and daily behavior of the runoff of the Xallas river using Box-Jenkins and neural networks methods *Journal of Hydrology*. 296. 38-58.
3. Cryer, D.j. and K.C. Chan. 2008. *Time Series Analysis with Applications in R*. Springer Texts in Statistics. Second Edition. 501pp.
4. Dong, X., F.M. Dohmen-Janssen, M. Booij and S. Hulscher. 2006. "Effect of flow forecasting quality on benefits of reservoir operation – a case study for the Geheyan reservoir (China)." *Hydrology and Earth System Sciences Discussions*.
5. Noakes, D.J., A.I. McLeod and K.W. Hipel. 1985. "Forecasting monthly riverflow time series." *International Journal of Forecasting*, 1, 179-190.
6. Parker, D., S. Tunstall and T. Wilson. 2005. "Socio-economic benefits of flood forecasting and warning." In *Proceedings of the International Conference on Innovation, Advances and Implementation of Flood Forecasting Technology*, October 17-19, 2005, Norway, Tromso.
7. Salas, J.D., J.W. Delleur, V. Yevjevich and W.L. Lane. 1980. *Applied modeling of hydrologic time series*. Water Resource Publication. 484 p.
8. Salas, J.D. and J.T. Obeysekera. 1982. ARMA model identification of hydrologic time series. *Water Resour. Res.* 18(4): 1011-1021.
9. Shalamu, A. 2009. *Monthly and seasonal streamflow forecasting in the Rio Grande Basin*. Doctor of Philosophy dissertation, New Mexico State University Las Cruces, New Mexico.
10. Tawfik, M. 2003. Linearity versus non-linearity in forecasting Nile River flows. *Advances in Engineering Software* 34. PP. 515-524.
11. Venables, W.N. and D.M. Smith. 2011. *An Introduction to R*. ISBN 3-900051-12-7.
12. Wang, W., P.H.A.J.M. van Gelder and J.K. Vrijling. 2005. Constructing prediction interval for monthly streamflow forecasts. In J. K. Vrijling et al. (Eds.) *Proceedings of the 9th International Symposium on Stochastic Hydraulics*, May 23- 24, 2005, Nijmegen, Netherlands. International Association of Hydraulic Research, Madrid, Spain.
13. Wang, W., P.H.A.J.M. Van-Gelder, J.K. Vrijling and J. Ma. 2006. Forecasting daily streamflow using hybrid ANN models. *Journal of Hydrology* 324 (2006) 383-399.

## Analysis Ability of the Autoregressive Integrated Moving Average Model for Forecasting of Reservoir Daily Inflow of Dez Reservoir with Two-year lead Time

M. E. Banihabib<sup>1</sup>, R. Bandari<sup>2</sup> and S. S. Mosavi Nadoshani<sup>3</sup>

### Abstract

One of the usual methods for stream flow prediction are time series models. In this research, an autoregressive integrated moving average model (ARIMA) is used for forecasting of daily inflow of Talezang station, located in upstream of Dez dam. As the data have seasonal trend, statistical indices (mean factors and standard deviation) of daily discharge, are estimated for 28 years with period of 365 days using Fourier series. Then, daily-observed discharge data were standardized using the statistical indices. The results of the research showed that seasonal trend of data was removed by the calculated factors of Fourier series. Then, different autoregressive integrated moving average models were fit into standardized data. Finally, by Akaike Information Criterion (AIC) and considering the minimum number of factors of model, the best model was selected. The results of forecasting by selected model showed that the model can forecast daily inflow trend relatively suitable and comparing with previous researches, mean absolute relative error of daily flow forecasting decreased from 3.12 to 0.6 and forecast lead-time increased from 10 days to two years.

**Keywords:** ARIMA, Daily inflow forecast, Dez Dam, Fourier series, seasonal tend of data.

---

<sup>1</sup> PhD, Assistant Professor, Department of Irrigation and Drainage Engineering, University of Tehran, (Banihabib@ut.ac.ir)

<sup>2</sup> M.Sc. Student, Water Resources Engineering, University of Tehran, Iran. (R\_bandary@ut.ac.ir)

<sup>3</sup> PhD, Assistant Professor, Department of water Engineering, University of water and power technology (Abaspour)(s\_mousavi@pwut.ac.ir)