

# کاربرد مدل‌های آماری بیزی در تحلیل داده‌های فیزیکی و شیمیایی آب‌های سطحی شمال و جنوب شرق تهران

غلامرضا اسدالله‌فردی (استادیار)

دانشکده مهندسی عمران و محیط زیست، دانشگاه تربیت معلم

سیدياسر صمدی (دانشجوی دکتری)

محمدرضا مشکانی (استاد)

گروه آمار، دانشگاه شهید بهشتی

با توجه به اینکه ایران جزو مناطق کم آب دنیا است، استفاده‌ی بهینه و جلوگیری از آلودگی آب امری ضروری است. این امر بدون داشتن اطلاعات از وضعیت کیفی آب امکان‌پذیر نیست. در این تحقیق به‌عنوان مطالعه‌ی موردی پیرامون شبکه‌ی آب‌های سطحی تهران، شاخه‌ی شمال شرقی و جنوب در دو ایستگاه در شمال شبکه و جنوب آن، تعدادی از پارامترهای فیزیکی و شیمیایی مورد پایش قرار گرفته‌اند. با توجه به این که فواصل زمانی پایش از نظم خاصی پیروی نمی‌کرد، برای تحلیل و تفسیر داده‌ها از مدل‌های سری زمانی بیزی استفاده شده است. نتایج تحلیل داده‌ها نشان‌گر برازش مناسب مدل رگرسیون پویا بر پارامترهای مواد محلول و اکسیژن محلول در آب شبکه است، و پیش‌بینی مطمئن آن را به دنبال داشت. برای اسیدیته‌ی آب چندجمله‌یی درجه دو برازش بهتری را نشان می‌دهد. در نتیجه مدل‌های بیزی را می‌توان به‌عنوان ابزار مناسبی در تحلیل متغیرهای کیفی آب به‌کار برد؛ و از آن در پیش‌بینی کوتاه‌مدت متغیرهای کیفی آب‌های سطحی در مدیریت کنترل کیفی منابع آب استفاده کرد.

fardi@tmu.ac.ir  
s.y.samadi@gmail.com  
mrmeshkani@gmail.com

واژگان کلیدی: پایش کیفی آب‌های سطحی، مدل‌های بیزی، مدل‌های رگرسیونی پویا.

## مقدمه

پیش‌بینی کیفیت منابع آب در چند دهه‌ی اخیر مورد توجه جدی صاحب‌نظران قرار گرفته است.

اولین بار تومن<sup>[۱]</sup> برای تحلیل منابع آب از سری زمانی به‌منظور تغییرات درجه و مواد محلول در خلیج دلاویر استفاده کرد. داده‌های مورد استفاده‌ی تومن داده‌های پایش پیوسته‌ی ایستگاه‌های اندازه‌گیری بودند که توسط مرکز جغرافیای آمریکا و شهر فیلادلفیا پایش شده بودند. پس از آن محققین گزارش موفقیت‌آمیزی از کاربرد سری زمانی باکس و جنکینز ارائه کردند.<sup>[۲]</sup> هاگ و فارکوهر<sup>[۳]</sup> از سری زمانی باکس و جنکینز<sup>[۴]</sup> برای دو متغیر کلرواکسیژن محلول در رودخانه‌ی کلرتردیک کرونا در ایالت انتاریو استفاده کردند و نتیجه‌ی آن موفقیت‌آمیز بود. لوهانی و ونگ<sup>[۵]</sup> نیز از سری زمانی در داده‌های ماهانه‌ی متغیرهای کیفی آب برای رودخانه‌ی چانگ کنگ واقع در قسمت شمالی استان در میانه‌ی تایوان استفاده کردند. از سری زمانی ARIMA به‌منظور پیش‌بینی کیفیت آب استفاده شد.<sup>[۶]</sup> همچنین برای تحلیل کیفیت آب رودخانه‌ی کلمیا Revelstoke، سری زمانی باکس و جنکینز مورد استفاده قرار

با توجه به افزایش جمعیت جهان، بالارفتن بهداشت عمومی و زیرکشت بردن زمین‌های بیشتر برای تولیدات کشاورزی و توسعه‌ی صنعت، مصرف آب در جهان افزایش یافته است. لذا مسئله‌ی آب به چالش و مسئله‌ی اصلی بعضی از کشورها تبدیل شده است. همچنین با افزایش جمعیت، پیشرفت‌های صنعتی، کشاورزی و بهداشت عمومی در ایران نیز، مصرف آب در کشورمان افزایش یافته است. از سوی دیگر به‌دلیل قرارگرفتن ایران در منطقه‌ی نیمه‌خشک و توزیع نامتناسب آب موجود در کشور، با مسئله‌ی کم‌آبی مواجه هستیم. لذا مدیریت صحیح منابع آب، داشتن اطلاعات کافی و به‌روز از کمیت و کیفیت در برنامه‌ریزی، و مدیریت منابع آبریزی مهم و ضروری است. داشتن اطلاعات فراوان از آب به‌تنهایی کارگشا نیست بلکه استفاده از روند تغییرات کیفی در هرگونه برنامه‌ریزی مدیریتی به‌منظور کنترل منابع مستلزم تحلیل و تفسیر داده‌های کیفی است. استفاده از معادلات سری زمانی به‌منظور

تاریخ: دریافت ۱۳۸۶/۱۱/۳، داوری ۱۳۸۷/۴/۹، پذیرش ۱۳۸۷/۱۰/۲.

گرفت.<sup>[۹]</sup> پس از آن پژوهش‌گران با استفاده از مدل‌های تحلیل رگرسیون چندمتغیره و نیز از سری زمانی باکس جنکینز به مطالعه‌ی درجه حرارت در مسیل Catamaran Brook پرداختند.<sup>[۱۰]</sup> همچنین می‌توان به تجزیه و تحلیل داده‌های کیفی آب توسط محققین نیز اشاره کرد.<sup>[۱۱]</sup> در ایران نیز سری‌های زمانی داده‌های ماهانه‌ی کیفیت آب‌های سطحی شهر تهران با استفاده از روش باکس جنکینز و هموارسازی نمایی<sup>۱</sup> تحلیل شده است.<sup>[۱۲]</sup> برای انجام این آنالیز طی مدت سه سال ۱۵ آلاینده‌ی آب‌های سطحی شهر تهران را از ۱۱ ایستگاه پایش کردند و در نهایت به مدل‌سازی و بحث در مورد آنها پرداختند؛ بیشتر مدل‌های حاصل خصوصیات فصلی از خود نشان دادند. در سال ۲۰۰۵ داده‌های کیفی آب به دو روش ARIMA و Thomas-Fiering مدل‌سازی شد.<sup>[۱۳]</sup> پژوهش‌گران این مطالعه را بر روی رودخانه‌ی Yesilirmark در ایستگاه دراکاسو انجام دادند. به این منظور داده‌های سری زمانی ماهانه به مدت ۱۳ سال استفاده شد تا بهترین مدل ممکن با هر دو روش به دست آید.

در این تحقیق به مدل‌سازی پارامترهای کیفی آب‌های سطحی شمال و شرق تهران که از شمال تهران به سمت جنوب شرقی تهران در جریان‌اند پرداخته‌ایم. در این مطالعه دو ایستگاه یکی در بالادست شبکه، و دیگری در پایین دست آن انتخاب شد (شکل ۱ الف). وسعت حوزه‌ی مورد مطالعه حدود ۴۷۵ کیلومتر مربع است و کاربری آن عمدتاً به‌عنوان مسکونی، اداری، تجاری و صنعتی است. در جنوب حوزه کارگاه‌های صنعتی قدیمی مثل سنگ‌بری، قالی‌شویی و آب‌کاری قرار دارند. کاربری اصلی شبکه‌ی مورد مطالعه جمع‌آوری آب‌های سطحی است. علاوه بر روان‌آب‌های سطحی، پساب خانگی بعضی ساختمان‌ها و پساب تصفیه‌خانه‌های

قیطریه، صاحب‌خانه، شوش و دولت‌آباد که بعضاً خروجی آنها به دقت کنترل نمی‌شود، وارد این مسیل‌ها می‌شود. همچنین ورود پساب کارگاه‌های صنعتی ذکرشده از آلاینده‌های عمده‌ی شبکه‌اند که با ورود آنها شبکه‌ی مورد مطالعه کاربری اصلی خود را از دست داده است.

هدف از این مطالعه عبارت است از: ۱. پایش تعدادی از متغیرهای کیفی شامل اکسیژن محلول، کل مواد محلول، اسیدیته، هدایت الکتریکی و درجه حرارت آب؛ ۲. پیدا کردن مدل مناسب و مطمئن سری زمانی پیزی برای پارامترهای مذکور در ایستگاه بالا دست و پایین دست؛ ۳. پیش‌بینی آینده‌ی روند تغییرات کیفی با استفاده از مدل سری زمانی، به‌منظور استفاده از نتایج تحلیل در مدیریت کیفیت منابع آب.

## نظریه و روش تحلیل داده‌ها

### ۱. مدل‌های خطی پویا

شیوه‌ی مدل‌سازی آماری مورد نظر برای فرایندهای مولد سری‌های زمانی، براساس رده‌ی مدل‌های پویاست. واژه‌ی پویا برای بیان تغییراتی است که بر اثر گذشت زمان در ساختار این‌گونه سری‌های پدید می‌آیند. رده‌ی بسیار شناخته‌شده‌ی این مدل‌ها که کاربرد وسیعی هم دارد، مدل‌های خطی پویای<sup>۲</sup> نرمال است. در پیش‌بینی مقادیر آتی سری برای هر نقطه‌ی زمانی  $t > 0$  هدف اصلی، محاسبه‌ی توزیع پیش‌بینی  $(Y_t | D_0)$  است. حال اگر  $D_t$  مجموعه‌ی اطلاعات در زمان باشد، آنگاه برای پیش‌بینی مقدار آینده سری برای زمان‌های  $s > t$  باید توزیع پیش‌بینی  $(Y_s | D_t)$  مورد توجه قرار گیرد. مشخص است که با گذشت زمان اطلاعات اولیه‌ی پیش‌بینی‌کننده بیشتر می‌شود و مجموعه‌های اطلاعاتی  $D_t$  گسترش می‌یابد.<sup>[۱۴، ۱۵]</sup>

به‌طور کلی در مدل‌های آماری مشاهدات مستقل از هم و با توزیع نرمال یکسان فرض می‌شوند، یعنی به‌ازای  $t = 1, 2, \dots$  خواهیم داشت:  $(Y_t | \mu) \sim N(\mu, V)$  وقتی از سیستم یا فرایندی که خود دائماً در حال تغییر است مشاهداتی به دست می‌آید، وقوع تغییرات در میانگین این مدل (و گاهی اوقات واریانس آن) با گذشت زمان طبیعی و اجتناب‌ناپذیر است. اگرچه تغییر برحسب زمان می‌تواند آهسته و تدریجی باشد، برای رفع مشکل این الگوی ساده و انعطاف‌پذیر کافی است اجازه دهیم مثلاً میانگین با زمان تغییر کند. چنین مدل‌هایی در پیش‌بینی‌های کوتاه‌مدت کاربرد زیادی دارند.<sup>[۱۵]</sup>

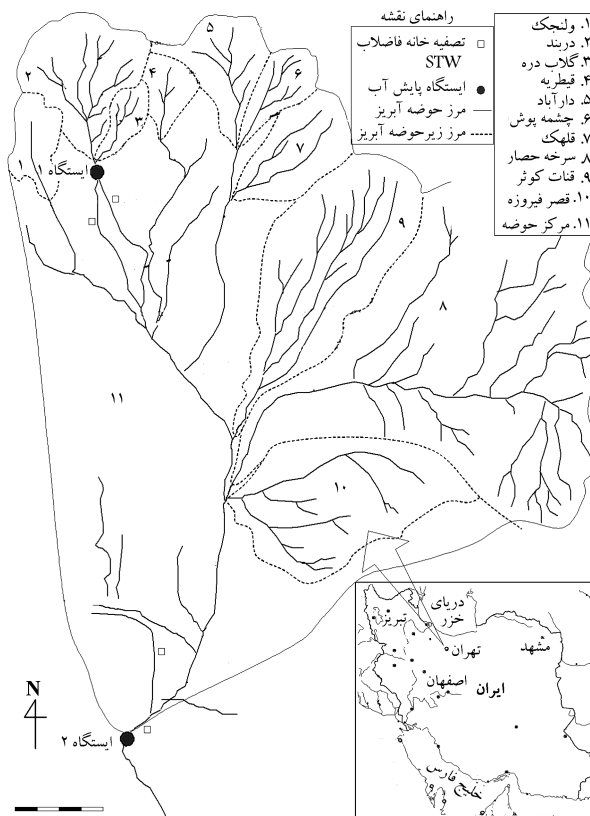
مدل خطی پویای نرمال برای متغیر  $Y_t$  را با چهارگانه‌ی  $\{F, G, V, W\}_t$  نشان می‌دهند. این چهارگانه معرّف مدلی است که  $Y_t$  را به بردار پارامتر  $\theta_t$  مربوط می‌سازد و برحسب زمان از طریق توزیع‌های دنباله‌ی زیر (رابطه‌های ۱ تا ۳) تعریف می‌شود:<sup>[۱۶]</sup>

$$Y_t = F_t^T \theta_t + v_t, \quad v_t \sim N(0, V_t) \quad (1)$$

$$\theta_t = G_t \theta_{t-1} + \omega_t, \quad \omega_t \sim N[0, W_t] \quad (2)$$

$$(\theta_0 | D_0) \sim N[m_0, C_0] \quad (3)$$

به معادلات ۱ تا ۳ به ترتیب معادله‌ی مشاهده، معادله‌ی تحول<sup>۳</sup> و توزیع پیشین اولیه گفته می‌شود که در آنها  $m_0$  و  $C_0$  مقادیر معلوم‌اند. معادله‌ی مشاهده، مشاهدات را به‌وسیله‌ی رگرسیون خطی پویا و خطای نرمال چندمتغیره‌ی بی‌با ساختار معلوم و متغیر نسبت به زمان، به  $\theta_t$  مرتبط می‌سازد. ماتریس  $F_t$  نقش ماتریس طرح رگرسیونی را ایفا می‌کند و حاوی مقادیر معلوم برای بردارهای مستقل است. بردار  $\theta_t$  پارامترهای رگرسیونی پویا یا «بردار وضعیت مدل» است. عبارت  $\mu_t = F_t^T \theta_t$  تابع میانگین



شکل ۱ الف. موقعیت حوزه‌ی مورد مطالعه و ایستگاه‌های اندازه‌گیری پارامترهای کیفی آب.

$$\begin{aligned} (\mu_{t-1} | D_{t-1}) &\sim N(m_{t-1}, C_{t-1}) && \text{الف) توزیع پسین برای } \mu_{t-1} \\ (\mu_t | D_{t-1}) &\sim N(m_{t-1}, R_t) && \text{ب) توزیع پیشین برای } \mu_t \\ (Y_t | D_{t-1}) &\sim N(f_t, Q_t) && \text{ج) توزیع پیش‌بینی یک گام به جلو} \\ (\mu_t | D_{t-1}) &\sim N(m_t, C_t) && \text{د) توزیع پسین برای } \mu_t \end{aligned}$$

که در آنها

$$\begin{aligned} e_t = Y_t - f_t, \quad f_t = m_{t-1}, \quad Q_t = R_t + V_t, \quad A_t = \frac{R_t}{Q_t} \\ C_t = A_t V_t, \quad m_t = m_{t-1} + A_t e_t, \quad R_t = C_{t-1} + W_t \end{aligned} \quad (7)$$

به عبارت دیگر  $A_t$  ضریب رگرسیونی و مجذور ضریب همبستگی است. همچنین می‌توان شکل دیگری برای  $m_t$  به صورت  $m_t = A_t Y_t + (1 - A_t) m_{t-1}$  در نظر گرفت؛ یعنی  $m_t$  میانگین موزونی از برآورد پیشین سطح  $m_{t-1}$  و مشاهده‌ی  $Y_t$  است. از این رو  $A_t$  را «ضریب سازواری» نیز می‌نامند. باید توجه داشت وقتی که واریانس مشاهداتی و واریانس تحول نامعلوم‌اند، آنگاه رابطه‌ی ۷ با رابطه‌ی ۸ تکمیل می‌شود که از آنها می‌توان در تحلیل مانده‌ها استفاده کرد.

$$\begin{aligned} C_t = \frac{S_t}{S_{t-1}} (R_t - A_t' Q_t), \quad W_t = C_{t-1} \left( \frac{1 - \delta_t}{\delta_t} \right), \\ d_t = d_{t-1} + S_{t-1} \frac{e_t'}{Q_t}, \quad d_{t-1} = S_{t-1} \times n_{t-1}, \quad S_t = \frac{d_t}{n_t}, \\ u_t = \frac{e_t}{\sqrt{Q_t}}, \quad Q_t = R_t + S_{t-1} \end{aligned} \quad (8)$$

ضرایب تخفیف اِبراری برای انتخاب  $W_t$  هستند. طبق تعریف مدل ثابت داریم:  $W = \left( \frac{A}{1-A} \right) C$  در حالت حدی  $R_t = C_{t-1} + W$  و بنابراین در دست می‌آید، یعنی  $W$  مضرب ثابتی از  $C$  است.<sup>[۱۶]</sup> این یک شیوه‌ی طبیعی نگرش در مورد واریانس تحول است. در رابطه‌ی ۸ پارامتر  $\delta$  را که مقدار آن برابر  $1-A$  است «ضریب تخفیف» می‌نامند. پس با انتخاب  $\delta$  می‌توان  $W$  را براساس ضریب سازواری حدی از روی داده‌ها محاسبه کرد. به ازای  $A = 0/1$  مقدار  $\delta$  برابر  $0/9$  خواهد بود و تقریباً ۱۱ درصد  $C$  خواهد بود. این میزان با انتخاب  $\delta = 0/8$  به ۲۵ درصد می‌رسد. بنابراین بهتر است که ضریب تخفیف ( $\delta$ ) اغلب  $0/8$  یا  $0/9$  در نظر گرفته شود. واضح است که این مدل ثابت نیست بلکه در حد، مدل ثابت با  $A = 1 - \delta$  است.

#### ۴. مدل پویای خطی رگرسیونی چندگانه

مدل بندی سری  $Y_t$  را با تعیین رگرسیون آن برحسب  $n$  متغیر مستقل به صورت  $X_{t1}, \dots, X_{tn}$  در نظر بگیرد ( $t = 1, \dots, n$  و  $i = 1, \dots, n$ ). در این صورت مدل پویای خطی رگرسیونی چندگانه با متغیرهای رگرسیونی  $X_1, \dots, X_n$  با چهارگانه‌ی  $\{F_t, I, V_t, W_t\}$  نشان داده می‌شود.<sup>[۱۷]</sup> در نتیجه به ازای هر  $t$ ، مؤلفه‌های معادلات ۱ و ۲ را می‌توان چنین نوشت:

$$F_t' = (X_{t1}, \dots, X_{tn}), \quad G_t = G = I, \quad \theta_t' = (\theta_{t1}, \dots, \theta_{tn})$$

در نتیجه می‌توان معادله‌ی مشاهده را به صورت  $Y_t = \mu_t + v_t$  نوشت که در آن  $\mu_t$  میانگین پاسخ به صورت  $\mu_t = F_t' \theta_t = \sum_{i=1}^n \theta_{ti} X_{ti}$  تعریف می‌شود. این مدل یک تعمیم پویا از مدل رگرسیونی ایستای استاندارد است. با قرار دادن  $W_t = 0$  به ازای هر  $t$ ، این مدل به حالت خاص مدل رگرسیونی ایستا تبدیل می‌شود؛ زیرا در این صورت  $\theta_t = \theta$  در طول زمان ثابت می‌ماند.<sup>[۱۶]</sup>

پاسخ در زمان است که سطح سری را در زمان  $t$  نشان می‌دهد. بالاخره در معادله‌ی مشاهده،  $v_t$  خطای مشاهداتی در زمان  $t$  است.

معادله‌ی تحول، تحول بردار پارامتر سری را نسبت به زمان نشان می‌دهد. مؤلفه‌ی قطعی یا سیستماتیک این معادله، تحول یا انتقال از  $\theta_{t-1}$  به  $\theta_t$  است که از طریق «ماتریس انتقال تحول»  $G_t$  فراهم می‌شود. این معادله با افزودن جمله‌ی خطای تحول  $w_t$  که دارای واریانس  $W_t$  است، تکمیل می‌شود. دنباله‌ی خطای مشاهداتی و نیز دنباله‌ی خطای تحول مستقل از هم فرض می‌شوند. همچنین فرض می‌شود که این دو دنباله به طور متقابل مستقل از هم و مستقل از  $(\theta_e | D_e)$  هستند.

#### ۲. مدل چندجمله‌یی مرتبه‌ی اول

ساده‌ترین و پر مصرف‌ترین مدل خطی پویا مدل چندجمله‌یی مرتبه‌ی اول<sup>۴</sup> است. نمایش مدل چندجمله‌یی مرتبه‌ی اول در قالب مدل خطی پویای کلی چنین است:

$$\begin{aligned} Y_t = Y_t \quad F_t = F = 1 \quad V_t = V_t \quad w_t = w_t \\ \theta_t = \theta_t = \mu_t \quad G_t = G = 1 \quad W_t = W_t \quad v_t = v_t \end{aligned} \quad (4)$$

ویژگی مهم مدل فوق این است که اگر در زمان  $t$ ، تعداد  $k$  گام به سوی آینده پیش برویم، مقدار متوسط این سری‌های در آن زمان، به شرط دانستن سطح فعلی، عبارت است از:  $E[Y_{t+k} | \mu_t] = \mu_t$  توزیعی که شخص پیش‌بینی‌کننده برای  $(\mu_t | D_t)$  در نظر می‌گیرد، در عمل دارای میانگین  $m_t$  است که به اطلاعات گذشته بستگی دارد. بنابراین تابع پیش‌گوی  $f_t(k)$  به ازای هر  $(k, 0)$  به صورت رابطه‌ی ۵ نشان داده می‌شود. یعنی متوسط سطح آبی سری نسبت به زمان حال ثابت است. بنابراین از چنین مدلی فقط می‌توان برای پیش‌بینی‌های کوتاه مدت استفاده کرد، به ویژه در مواردی که تغییرات نمونه‌گیری ( $V_t$ ) در مقایسه با واریانس سطح ( $W_t$ ) به مقدار قابل توجهی بزرگ باشد.

$$f_t(k) = E[Y_{t+k} | D_t] = E[\mu_{t+k} | D_t] = E[\mu_t | D_t] = m_t \quad (5)$$

#### ۳. مدل چندجمله‌یی مرتبه‌ی دوم

مدل خطی پویای چندجمله‌یی مرتبه‌ی دوم، «مدل رشد خطی»<sup>۶</sup> در هر زمان  $t$  و  $k \geq 0$ ، دارای تابع پیش‌گویی به صورت خط راست  $(f_t(k) = a_{t0} + a_{t1}k)$  است. نمایش دیگر از رابطه‌ی بالا عبارت است از:  $f_t(k) = f_t(0) + [f_t(1) - f_t(0)]k$ ، برحسب دو مقدار نخست از تابع پیش‌گویی تعیین می‌شوند. نمایش معمول مدل خطی پویای چندجمله‌یی مرتبه‌ی دوم چنین است:

$$\begin{aligned} Y_t = \mu_t + v_t, \quad v_t \sim N(0, V_t) \\ \mu_t = \mu_{t-1} + \beta_t + w_{t1}, \quad w_{t1} \sim N(0, W_t(\mu)) \\ \beta_t = \beta_{t-1} + w_{t2}, \quad w_{t2} \sim N(0, W_t(\beta)) \end{aligned} \quad (6)$$

در نتیجه مؤلفه‌های معادلات ۱ و ۲ را می‌توان همانند رابطه‌ی ۴ در نظر گرفت.<sup>[۱۶]</sup> حال ساختار و خواص مدل چندجمله‌یی مرتبه‌ی اول را با هدف آشنایی با طبیعت دنباله‌یی معادلات به روزکننده و ربط دادن این مدل با سایر روش‌ها و مفاهیم پیش‌بینی متداول، بررسی می‌کنیم. برای هر  $t$  توزیع‌های پسین، پیشین و پیش‌بینی یک گام به جلو عبارت‌اند از:

حالت مهمی از مدل رگرسیونی عبارت است از یک رگرسیون خط راست برحسب  $X = X_t$ ، که به‌وسیله‌ی  $F_t = (\mathbf{1}, X_t)'$  و  $\theta_t = (\alpha_t, \beta_t)'$  فراهم می‌شود. بنابراین،

$$Y_t = \alpha_t + \beta_t X_t + v_t, \quad v_t \sim N(0, V_t) \quad (9)$$

$$\alpha_t = \alpha_{t-1} + \omega_{t1}, \quad \beta_t = \beta_{t-1} + \omega_{t2} \Rightarrow \omega_t = (\omega_{t1}, \omega_{t2})' \sim N(0, W_t) \quad (10)$$

## روش انجام کار و تحلیل

قسمت اعظم آب‌های سطحی تهران از نقاط کوهستانی شمال تهران سرچشمه می‌گیرد و به سمت جنوب شرق تهران جریان پیدا می‌کند. به علت وجود مراکز تفریحی و عبور از میان بافت مسکونی و تجاری - با توجه به عدم وجود سیستم جمع‌آوری و تصفیه‌ی فاضلاب - قسمتی از فاضلاب‌های خانگی-تجاری مستقیماً وارد این شبکه می‌شود و آب‌های سطحی را آلوده می‌سازد. در نهایت این آب در پایین دست برای کشاورزی مصرف می‌شود که تهدیدکننده‌ی سلامت عمومی است. همچنین به علت بالا رفتن املاح محلول ( $TDS$ ) ممکن است مصرف آن با مشکل مواجه باشد. در واقع این شبکه‌ی رودخانه‌ی براثر ورود پساب‌های خانگی، و گاهی تجاری، آلوده می‌شود.

با توجه به اهمیت این آب‌ها از نظر بهداشتی و نیز از نظر اقتصادی، در این تحقیق برآنیم که نسبت به مدل‌بندی متغیرهای شاخص کیفیت آب در این حوزه از دیدگاه سری‌های زمانی پویا پردازیم. شایان ذکر است که در این کاربرد، نمونه‌گیری‌ها در فاصله‌های زمانی برابر صورت نگرفته‌اند. بنابراین امکان استفاده‌ی مستقیم از سری‌های زمانی عادی در اینجا میسر نیست. اما چون مدل‌های پویای پیزی چنین محدودیتی ندارند، به تحلیل مسئله از دیدگاه این مدل‌ها می‌پردازیم.

در این راستا به منظور بررسی کیفیت آب‌های سطحی شمال و جنوب شرق تهران، نمونه‌برداری از برخی متغیرهای کیفی آب در دو ایستگاه از شبکه‌ی مذکور انجام شد. این دو ایستگاه یکی در بالادست این شبکه (یعنی «ایستگاه پل‌رومی») واقع در تقاطع خیابان قیصریه و شریعتی (و دیگری در پایین دست حوزه‌ی آبریز مورد مطالعه‌ی «ایستگاه صفائی») شهر ری، نزدیک خیابان ۲۵ متری، کنار کارخانه‌ی سیمان است. این ایستگاه‌ها در نقشه‌ی حوزه‌ی آبریز مشخص شده‌اند. آب از نقاط شمالی با کیفیت نسبتاً مطلوب به سمت جنوب حرکت می‌کند و به علت استقرار مناطق مسکونی، رستوران‌ها و بعضی کارگاه‌ها در اطراف آن، به تدریج آلوده می‌شود. بنابراین ایستگاه بالایی تقریباً در بالاترین نقطه‌ی حوزه و ایستگاه بعدی در پایین‌ترین

نقطه‌ی آن انتخاب شده است تا مقایسه‌ی تفاوت کیفیت آب در بالادست با کیفیت آن در پایین دست، پس از ورود پساب‌های شهری و صنعتی به این رودخانه، میسر باشد. هدف این بررسی مدل‌بندی متغیرهای شاخص کیفیت آب با استفاده از مدل‌های پویای پیزی سری زمانی است.

متغیرهای اندازه‌گیری شده در دو ایستگاه یادشده عبارت‌اند از  $EC, DO, pH$ ،  $T, TDS$  که از متغیرهای مهم فیزیکی و شیمیایی در کیفیت آب هستند. این متغیرها به مدت شش ماه و به تعداد ۴۵ مرتبه (از تاریخ ۸۳/۹/۱۳ تا ۸۴/۳/۲۵) در هر ایستگاه توسط محققین اندازه‌گیری شده‌اند. به منظور بررسی اولیه و تعیین شاخص‌های مرکزی و پراکندگی داده‌های معرفی شده، خلاصه‌ی آماره‌های توصیفی متغیرهای اندازه‌گیری شده در جدول ۱ الف ارائه شده است. در تحلیل‌های ارائه شده از نرم‌افزار تحلیل سری زمانی پیزی<sup>۱۸</sup> استفاده شده است.

با توجه به داده‌های خام و جدول ۱ الف، ملاحظه می‌شود که میزان اکسیژن محلول از ایستگاه ۱ (ایستگاه شمالی) به سمت ایستگاه ۲ (ایستگاه جنوبی) کاهش می‌یابد، که دلیل آن ورود پساب‌های مختلفی است که به این مسیر (در حد فاصل دو ایستگاه) تخلیه می‌شود. همین وضعیت در مورد کل مواد محلول برقرار است و چون هدایت الکتریکی به شدت به کل مواد محلول در آب بستگی دارد، میزان آن نیز از ایستگاه شمالی به سمت ایستگاه جنوبی همانند کل مواد محلول افزایش می‌یابد. از طرفی  $pH$  آب در دو ایستگاه تفاوت چندانی ندارد و تقریباً ثابت است.

به علت محدودیت زمانی و کمبود امکانات، پایش متغیرها به مدت بیشتر برای نگارندگان میسر نبود ولی با همین تعداد داده، مدل‌های مناسب تعیین و صحت‌سنجی آن‌ها نیز بررسی شده است و برای پیش‌بینی‌های کوتاه‌مدت قابل اعتماد ارزیابی شدند.

## ساخت مدل برای داده‌های اصلی

### ۱. برازش یک مدل چندجمله‌ی مرتبه‌ی اول

در این قسمت به عنوان اولین گام سعی داریم یک مدل چندجمله‌ی مرتبه‌ی اول را با پارامتر سطح  $\mu_t$  به سری زمانی  $TDS \setminus$  برازش دهیم. براین اساس در حالتی که  $V_t$  ثابت و نامعلوم است، نمایش این مدل در قالب مدل خطی پویا چنین است:<sup>۱۹</sup>

$$TDS \setminus = \mu_t + v_t, \quad v_t \sim N(0, V) \quad (11)$$

$$\mu_t = \mu_{t-1} + \omega_t, \quad \omega_t \sim N(0, VW_t^* = W_t) \quad (12)$$

جدول ۱ الف. خلاصه‌ی آماره‌های توصیفی متغیرهای اندازه‌گیری شده در دو ایستگاه.

پارامتر متغیر	میانگین	انحراف معیار	کمینه	بیشینه	چارک اول	چارک دوم (میانانه)	چارک سوم
DO ۱ (mg/l)	۸٫۳۲	۰٫۸۷	۷	۹٫۸	۷٫۷۳	۸٫۱۴	۸٫۷۳
DO ۲ (mg/l)	۶٫۳۵	۰٫۹۲	۴٫۹	۸٫۴۷	۵٫۶۰	۶٫۲۳	۷٫۱۱
pH ۱	۷٫۸۷	۰٫۳۴	۶٫۷۹	۸٫۲۸	۷٫۷۵	۷٫۹۳	۸٫۱۲
pH ۲	۷٫۹	۰٫۳۷	۶٫۹۳	۸٫۷۲	۷٫۶۴	۷٫۹۴	۸٫۱۴
EC ۱ (mg/l)	۴۹۹	۱۳۲	۲۵۵	۷۰۵	۳۹۲	۴۷۳	۶۰۳
EC ۲ (mg/l)	۸۳۰	۷۱٫۹۴	۶۶۸	۱۰۱۹	۷۹۷	۸۲۹	۸۶۹
TDS ۱ (mg/l)	۲۵۸	۷۳٫۵۷	۱۲۷	۳۵۲	۱۹۵٫۸	۲۳۶	۳۲۵
TDS ۲ (mg/l)	۴۲۵	۴۲٫۵۲	۳۴۵	۵۵۸	۳۹۹	۴۲۹	۴۴۸

که فرض می‌شود:

$$(\mu_0 | D_0, V) \sim N(m_0, VC_0^* = C_0),$$

$$(\Phi | D_0) \sim G\left(\frac{n_0}{2}, \frac{d_0}{2}\right) \quad (13)$$

چون  $V$  نامعلوم است از پارامتر  $\frac{1}{V}$  به جای  $V$  و از  $(\Phi | D_0)$  به عنوان توزیع پیشین یا اطلاع اولیه استفاده شده است.  $m_0$  برآورد سطح اولیه و  $C_0$  اندازه‌ی عدم قطعیت در برآورد میانگین است. معمولاً واریانس مشاهدات ( $V_t$ ) بزرگ‌تر از واریانس تحول ( $W_t$ ) است. برای برازش مدل فرض می‌کنیم واریانس مشاهدات ثابت اما نامعلوم است ( $V_t = V$ ) و تدبیری که برای تشخیص واریانس تحول ( $W_t$ ) به کار می‌رود، با استفاده از ضریب تخفیف خواهد بود.

پس از تعیین  $F$  و  $G$ ، گام بعدی تعیین پیشین‌هاست که عبارت‌اند از پیشین برای پارامتر مدل و پیشین برای واریانس مشاهده‌ی. ارائه‌ی پیشین‌ها می‌تواند بر مبنای  $D_0$  باشد. چنین اطلاعاتی ممکن است شامل داده‌های گذشته (قبل از تاریخ ۱۳/۹/۸۳) مربوط به متغیرهای کیفی آب رودخانه، مدل‌ها، روابط یا تجربیات موجود در این زمینه باشد.

پیشین‌ها را با استفاده از پیشین‌های پیشنهادی نرم‌افزار تحلیل سری زمانی پی‌زی انتخاب می‌کنیم. برای راحتی در نوشتن قرار می‌دهیم  $TDS_1 = Y$ . آبر پارامترهای این پیشین‌ها عبارت‌اند از:

$$m_{0,1} = m_0 = \frac{Y_1 + Y_2}{2} = 355.5,$$

$$\sqrt{S_0} = \sqrt{\frac{(Y_1 - Y_2)^2}{2}} = 137.435, \quad \sqrt{C_0} = 307.04,$$

$$n_0 = 1, \quad \delta = 0.9$$

که در آن‌ها  $S_0$  برآورد اولیه  $V$  است. با توجه به آبر پارامترهای اولیه، می‌توان توزیع‌های پیشین زیر را نتیجه گرفت.

$$(\mu_0 | D_0) \sim T_1(m_0 = 355.5, C_0 = 902.52),$$

$$(\Phi | D_0) \sim G\left(\frac{n_0}{2} = \frac{1}{2}, \frac{d_0}{2} = \frac{902.52}{2}\right)$$

همان‌طور که پیش‌تر اشاره شد، از مهم‌ترین ویژگی‌های این مدل (طبق رابطه‌ی (5)) این است که مقدار مورد انتظار برای  $k$  گام به جلو همان سطح سری در زمان جاری است. در نتیجه از این مدل‌ها فقط در پیش‌بینی‌های کوتاه‌مدت استفاده می‌شود.

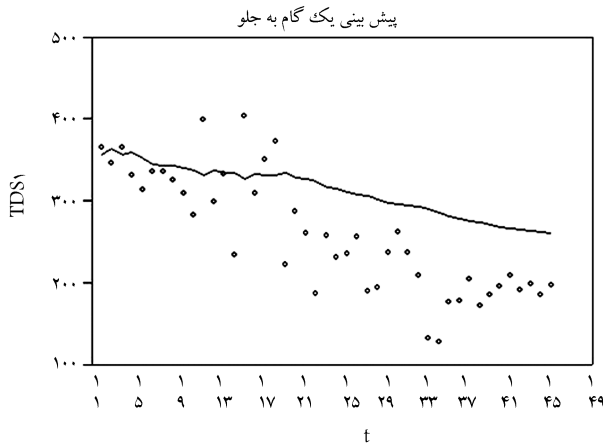
مدل چندجمله‌ی را می‌توان در دو حالت پویا و ایستا ( $W_t = 0$ ) به داده‌ها برازش داد. برای تشخیص این که مدل باید به صورت پویا یا ایستا به داده‌ها برازش یابد، ابتدا مدل‌هایی با ضرایب تخفیف مختلف را که یکی از آنها مدل ایستاست (یعنی  $\delta_T = 1$ ) به داده‌ها برازش می‌دهیم. از بین مدل‌های برازش یافته مدلی را انتخاب می‌کنیم که دارای کم‌ترین  $MSE$  باشد. با توجه به جدول ۱ و نمودارهای شکل ۱ ب و شکل ۲، مدل خطی پویا را انتخاب می‌کنیم. ضریب تخفیف توسط مدل‌ساز به‌گونه‌ی انتخاب می‌شود که شرایط مناسب برای مدل حاصل شود. در حقیقت هرچه ضریب تخفیف مقدار کم‌تری بگیرد پویایی و انطباق پذیری مدل افزایش می‌یابد، ولی عدم قطعیت مدل نیز افزایش می‌یابد.

در تمام شکل‌ها محور افقی که با  $t$  نمایش داده شده است، بیانگر زمان پایش متغیرهای فیزیکی و شیمیایی اندازه‌گیری شده است و عدد ۱ نشانگر اولین روز

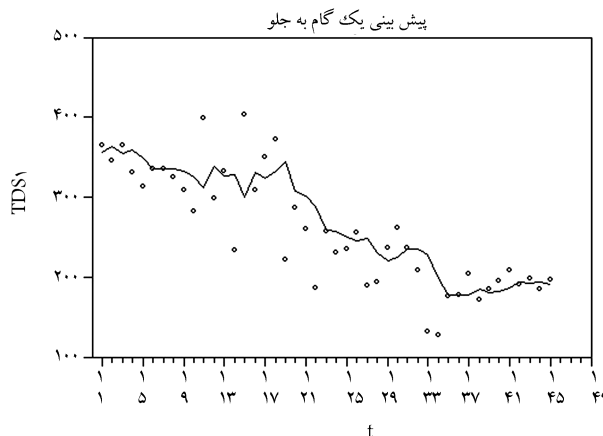
جدول ۱. خلاصه‌ی از برازش مدل چندجمله‌ی درجه اول به  $TDS_1$  با ضرایب تخفیف مختلف.

ضریب تخفیف	MSE	*MAD	log likelihood
$\delta_T = 1$	55622.4	63.73	-261.84
$\delta_T = 0.9$	2801.1	40.82	-246.35
$\delta_T = 0.8$	2111.4	32.95	-239.72
* $\delta_T = 0.7$	1967.4	30.78	-237.89
$\delta_T = 0.6$	1973.9	31.24	-237.72

\* میانگین قدر مطلق انحراف‌ها



شکل ۱. سری مشاهدات (نقطه‌ها) و سری مقادیر پیش‌بینی یک گام به جلو (خط پر) مدل چندجمله‌ی مرتبه‌ی اول از متغیر  $TDS_1$  با  $\delta_T = 1$ .



شکل ۲. سری مشاهدات (نقطه‌ها) و سری مقادیر پیش‌بینی یک گام به جلو (خط پر) مدل چندجمله‌ی مرتبه‌ی اول از متغیر  $TDS_1$  با  $\delta_T = 0.7$ .

نمونه‌گیری (یعنی تاریخ ۱۳/۹/۸۳) است. محورهای عمودی، نشانگر متغیرهای فیزیکی و شیمیایی و بعضاً مانده‌های استاندارد شده است که در شکل‌های مربوطه مشخص شده‌اند.

دلیل دیگر مناسب بودن مدل‌های پویا برای برازش به این داده‌ها، تغییراتی است که در سطح سری در طول زمان تحت بررسی رخ می‌دهد. در واقع متغیرهای آب رودخانه تحت تأثیر عوامل بیرونی زیادی قرار دارند. این‌گونه تغییرات را می‌توان

به‌وضوح در طول سری‌های مورد بحث دید. تا اینجا با توجه به جدول ۱ و معیار مدل پویا با ضریب تخفیف  $\delta_T = 0.7$  انتخاب شده است.

## ۲. برازش مدل چندجمله‌ی مرتبه‌ی دوم

به‌منظور بررسی بیشتر سری  $(TDS)$ ، تحلیل مدل چندجمله‌ی مرتبه‌ی دوم را با پارامترهای سطح و رشد  $(\beta_t, \mu_t)$  با استفاده از معادلات ۶ که پیش‌تر توضیح داده شد در نظر می‌گیریم. برای این مدل واریانس مشاهدات  $(V_t)$  به‌صورت پویا و نامعلوم در نظر گرفته می‌شود. ماتریس  $W_t$  نیز با استفاده از عامل تخفیف  $(\delta_T)$  مشخص می‌شود. برای انتخاب ضریب تخفیف مناسب باز مدل‌های مختلفی را به داده‌ها برازش می‌دهیم و از بین مدل‌های برازش داده شده مدلی را انتخاب می‌کنیم که دارای کم‌ترین  $MSE$  باشد. با توجه به جدول ۲ مدل پویایی با  $\delta_T = 0.9$  را انتخاب می‌کنیم.

## ۳. برازش مدل رگرسیونی

با توجه به هم‌بستگی زیاد  $TDS$  و  $T$  (دمای آب) می‌توان سری زمانی  $(T)$  را به‌عنوان سری زمانی مستقل و یک متغیر کمکی ورودی در این مدل در نظر گرفت. بنابراین سعی داریم مدل پویای خطی با اثرات رگرسیون را به داده‌ها برازش دهیم. برای راحتی در نوشتن  $TDS \setminus = Y$  و  $T \setminus = X$  فرض می‌کنیم. لذا با استفاده از معادلات ۹ و ۱۰ در مدل رگرسیونی خواهیم داشت:

$$(TDS \setminus)_t = \alpha_t + \beta_t(T \setminus)_t + v_t, \quad v_t \sim N(0, V_t)$$

واریانس مشاهداتی  $(V_t)$  همراه با پارامترهای دیگر مدل برآورد خواهد شد. واریانس عامل نوآوری تحول  $(W_t)$  با استفاده از تکنیک‌های تخفیف بلوکی مشخص می‌شود. چون دو مؤلفه‌ی بلوکی -- روند و متغیر رگرسیونی -- در مدل وجود دارد، ماتریس واریانس تحول عبارت خواهد بود از یک ماتریس قطری بلوکی با دو بلوک [۱۹، ۱۶] به‌خاطر داشته باشید که عامل‌های تخفیف کوچک‌تر به‌معنای عدم قطعیت بیشتر است. بنابراین همیشه سعی می‌شود عامل‌ها چنان انتخاب شوند که به هر آنچه لازم است بتواند تغییر یابد، به‌طوری که پویایی و تقریب مدل را به‌خوبی منعکس کند. سطح، در مدلی با مؤلفه‌ی رگرسیونی نقش عرض از مبدأ رگرسیون را ایفاء می‌کند. بنابراین بهتر است عامل تخفیف سطح و عامل تخفیف رگرسیونی نزدیک هم انتخاب شوند. برای انتخاب ضرایب تخفیف مناسب مدل‌های مختلفی را به داده‌ها برازش داده‌ایم (جدول ۳ الف). از بین مدل‌های برازش یافته مدلی را انتخاب می‌کنیم که دارای کم‌ترین  $MSE$  باشد.

با توجه به جدول ۳ الف و معیار  $MSE$ ، یک مدل پویا با ضرایب تخفیف  $\delta_T = 0.97$ ،  $\delta_R = 0.95$  را انتخاب می‌کنیم. از مقایسه‌ی

جدول ۲. خلاصه‌ی از برازش مدل چندجمله‌ی مرتبه‌ی دوم به  $TDS1$  با ضرایب تخفیف مختلف.

ضریب تخفیف	MSE	MAD	log likelihood
$\delta_T = 1$	۲۱۱۱٫۵	۳۵٫۴۸	-۲۴۱٫۹۷
$\delta_T = 0.95$	۲۱۰۲٫۵	۳۵٫۳۹	-۲۴۱٫۷۵
$\delta_T = 0.9$ *	۲۰۸۹٫۴	۳۴٫۸۷	-۲۴۱٫۴۵
$\delta_T = 0.85$	۲۰۹۷٫۱	۳۴٫۳۳	-۲۴۱٫۴۱
$\delta_T = 0.8$	۲۱۴۱٫۳	۳۳٫۹۸	-۲۴۱٫۸۰

$\delta_T$	$\delta_R$	۰٫۹	۰٫۹۵	۰٫۹۷	۰٫۹۹	۱
۰٫۹	MSE	۹۰۰٫۳	۹۵۳٫۵	۸۴۳٫۱	۸۷۵٫۵	۱۰۳۲٫۶
	MAD	۱۴٫۵۶	۱۵٫۱۱	۱۳٫۹۶	۱۴٫۳۱	۱۵٫۳۴
	log lik	-۱۵۶٫۸	-۱۵۷٫۳	-۱۵۵٫۶	-۱۵۶٫۹	-۱۵۶٫۸۶
۰٫۹۵	MSE	۸۶۶٫۷	۹۵۱٫۱	*۷۱۶٫۵۲	۸۱۶٫۲	۱۰۴۰٫۱
	MAD	۱۴٫۴۴	*۱۲٫۲۱	۱۳٫۲۴	۱۴٫۳۶	۱۵٫۷۵
	log lik	-۱۵۶٫۲	-۱۵۷٫۴	-۱۵۳٫۴	-۱۵۵٫۳	-۱۵۷٫۹
۰٫۹۸	MSE	۸۸۶٫۳	۹۵۲٫۵	۷۹۸٫۹	۸۵۱٫۷	۱۰۳۵٫۸
	MAD	۱۴٫۵۱	۱۵٫۱۲	۱۴٫۰۸	۱۴٫۱۶	۱۵٫۵۵
	log lik	-۱۵۶٫۵۶	-۱۵۷٫۲۵	-۱۵۵٫۱۲	*-۱۵۱٫۷	-۱۵۷٫۴
۱	MSE	۹۵۲٫۱	۹۷۴٫۷	۹۴۱٫۳	۹۴۶٫۱	۱۰۳۶٫۲
	MAD	۱۴٫۲۶	۱۴٫۳	۱۴٫۲۹	۱۴٫۲۹	۱۴٫۷۶
	log lik	-۱۵۶	-۱۵۶٫۳	-۱۵۶٫۷	-۱۵۷٫۱	-۱۵۹٫۱

جدول ۳ الف با جدول ۲ ملاحظه می‌شود که  $MAD$  و  $MSE$  در این حالت به‌طور چشم‌گیری کاهش یافته‌اند.

## ارزیابی برازش مدل به داده‌ها

در این قسمت به بررسی و ارزیابی مدل‌هایی می‌پردازیم که در قسمت قبل برازش داده شده‌اند، تا از این طریق بهترین مدل ممکن را انتخاب کنیم. تحلیل خطاها یا مانده‌ها یکی از بهترین روش‌ها برای ارزیابی و تشخیص مناسب بودن مدل است. با بهره‌گیری از معادلات روابط ۷ و ۸ می‌توان مقادیر مانده‌های خام و مانده‌های استاندارد شده را محاسبه کرد.

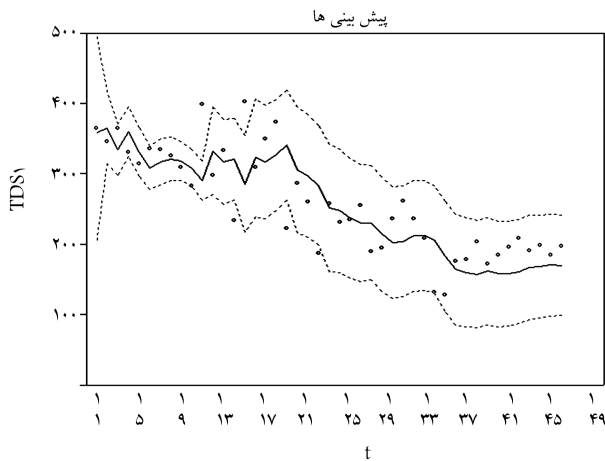
### ۱. ارزیابی برازش مدل چندجمله‌ی مرتبه‌ی اول به داده‌ها

خطاهای پیش‌بینی یک گام به جلو  $(e_t)$ ، ماده‌ی خام برای ارزیابی مدل است. بنابراین به‌منظور بررسی فرضیات مدل و ارزیابی مناسب بودن یا نبودن مدل عمدتاً از نمودارهای  $e_t$  یا مقدار استاندارد شده‌ی آن  $(u_t)$  استفاده می‌شود. در شکل ۳ نمودار مانده‌های استاندارد شده‌ی مدل چندجمله‌ی مرتبه‌ی اول نشان داده شده است که براساس  $\delta_T = \delta = 0.7$  محاسبه شده‌اند.

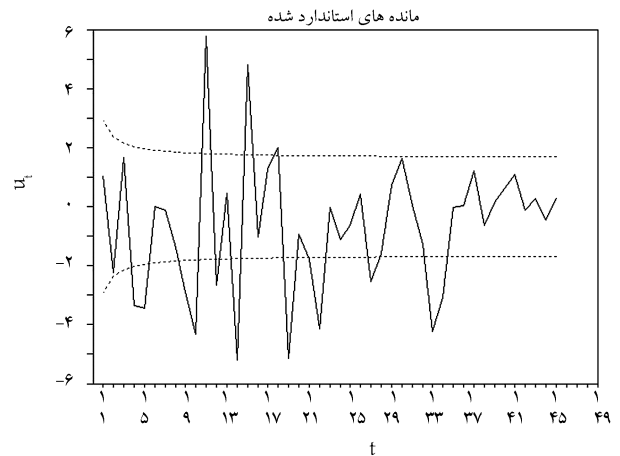
با نگاهی به این نمودار، به مجموعه‌ی از نقاط با قدر مطلق خطای زیاد، اما بدون الگو، برمی‌خوریم. علاوه بر این با توجه به مقادیر  $MSE = 2111.4$  و  $MAD = 32.95$  وجود نقاطی در خارج از بازه‌ی اطمینان در نمودار پیش‌بینی یک گام به جلو، در شکل ۴ نشانه‌هایی از عدم برازش مدل به داده‌ها موجود است.

### ۲. ارزیابی برازش مدل چندجمله‌ی مرتبه‌ی دوم به داده‌ها

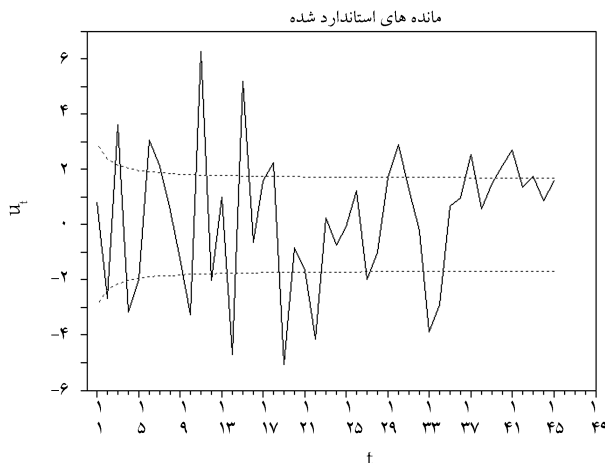
در مطالب پیشین نتیجه گرفتیم که مدل چندجمله‌ی مرتبه‌ی اول، مدل خوبی برای برازش داده‌های یاد شده نیست. اکنون به ارزیابی مدل چندجمله‌ی مرتبه‌ی دوم می‌پردازیم. در شکل ۵ نمودار پیش‌بینی یک گام به جلو و بازه‌های اطمینان ۹۵٪ با



شکل ۵. سری مشاهدات (نقطه‌ها) سری مقادیر پیش‌بینی یک گام به جلو (خط پر) مدل چندجمله‌ای مرتبه‌ی دوم از متغیر TDS1 با  $\delta_T = 0.9$  با بازه اطمینان ۹۵٪.



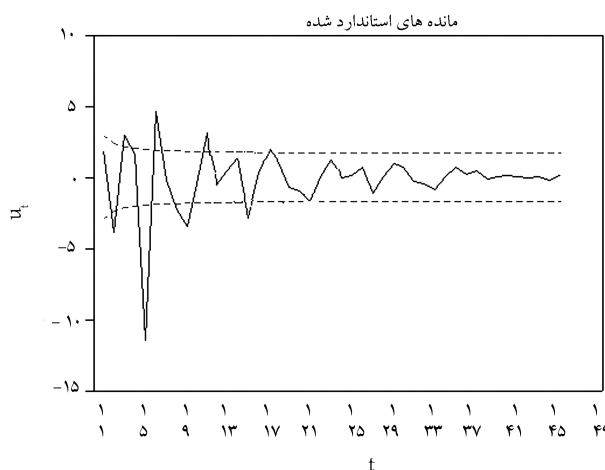
شکل ۳. سری مانده‌های استاندارد شده (خط پر) مدل چندجمله‌ای مرتبه‌ی اول از متغیر TDS1 با  $\delta_T = 0.7$  و با بازه اطمینان ۹۵٪ (خط چین).



شکل ۶. سری مانده‌های استاندارد شده مدل چندجمله‌ای مرتبه‌ی دوم از متغیر TDS1 با  $\delta_T = 0.9$ .



شکل ۴. سری مشاهدات (نقطه‌ها) و سری مقادیر پیش‌بینی یک گام به جلو (خط پر) مدل چندجمله‌ای مرتبه‌ی اول از متغیر TDS1 با  $\delta_T = 0.7$  با بازه اطمینان ۹۵٪ (خط چین).



شکل ۷. سری مانده‌های استاندارد شده مدل رگرسیونی از متغیر TDS1 با  $\delta_R = 0.95$  و  $\delta_T = 0.97$ .

$\delta_T = 0.9$  نشان داده شده است. وجود نقاطی در خارج از بازه اطمینان، افزایش تدریجی مقدار  $S_t$  و طولانی‌تر شدن بازه اطمینان مقادیر پیش‌بینی شده، نشانه‌هایی از عدم برازش مدل به داده‌ها است. شکل ۶ نمودار مانده‌های استاندارد شده را نشان می‌دهد که براساس  $\delta_T = 0.9$  محاسبه شده‌اند. اما باز هم مجموعه‌ی نقاط با قدر مطلق خطای زیاد داریم که این نیز گواهی دیگر بر عدم برازش مدل به داده‌ها است.

### ۳. ارزیابی برازش مدل رگرسیونی به داده‌ها

اکنون نشان می‌دهیم مدل رگرسیونی مدلی خوب و مناسب برای این داده‌ها است. بررسی گرافیکی یا عددی مانده‌ها از بهترین و مفیدترین ابزارها برای رسیدن به دیدگاهی درخصوص مناسب بودن یا نبودن مدل و همچنین شیوه‌ی بهسازی مدل است. در شکل ۷ نمودار مانده‌های استاندارد شده، که براساس  $\delta_R = 0.95$  و  $\delta_T = 0.97$  و با استفاده از نرم‌افزار BATS محاسبه شده‌اند نشان داده شده است.

### پیش‌بینی با رگرسیون

انجام پیش‌بینی از روی مدل‌هایی که شامل رگرسیون برحسب متغیرهای ظاهری‌اند، پیچیده‌تر از انجام پیش‌بینی از یک مدل ایستا باشد. در واقع مسئله‌ی مهم تعیین مقدار متغیرهای رگرسیونی در دوره‌ی است که پیش‌بینی صورت می‌گیرد. در وضعیت‌های واقعی، این مسئله یکی از مشکل‌ترین جنبه‌های پیش‌بینی است، زیرا مقادیر آینده‌ی متغیرهای ظاهری خودشان قطعیت ندارند. یک روش کنترل برای این اتفاق، محاسبه‌ی مجموعه‌هایی از پیش‌بینی‌ها برای تغییر وضعیت‌های ممکن، متناظر با پیشامدهای مختلف ممکن از متغیرهای ظاهری است.

در عمل، ممکن است پیش‌بینی‌کننده مدل‌های دیگری در اختیار داشته باشد که از آن برای پیش‌بینی ورودی‌های مورد نیاز (متغیرهای ظاهری) استفاده کند. البته این کار باید با احتیاط انجام شود. غالباً در مدل‌های پویای رگرسیونی، رگرسیون برحسب متغیرهای تبیینی نسبت به مؤلفه‌های سری زمانی محض مانند روندها، حرکت زمان‌مند کم‌تری را به نمایش می‌گذارند. به این دلیل ضرایب کاهش رگرسیونی مدل‌های اولیه بزرگ‌تر از ضرایب تخفیف روند در نظر گرفته می‌شود.

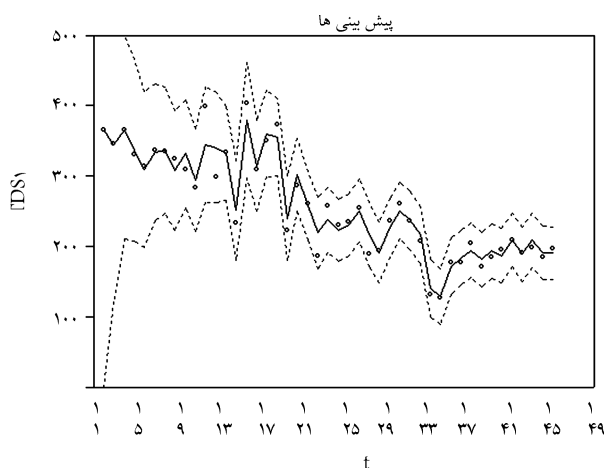
برای پیش‌بینی مقادیر آینده‌ی  $TDS_1$  نیاز به مقادیر آینده‌ی  $T_1$  داریم. چون مقادیر آینده‌ی  $T_1$  نامعلوم‌اند، در اینجا برای جایگزین کردن این مقادیر، روشی را به کار می‌گیریم که در آن آخرین مشاهدات ( $T_1$ ) به جای مقادیر آینده‌ی آن قرار می‌گیرد. بر این اساس پنج مقدار آخر، یعنی مشاهدات ۴۱ تا ۴۵ از  $T_1$  را به ترتیب در زمان‌های ۴۶ تا ۵۰ از  $T_1$  قرار می‌دهیم، تا بتوان مقادیر ۴۶ تا ۵۰ از  $TDS_1$  را پیش‌بینی کرد. توجه شود که از  $T_1$  فقط به عنوان یک متغیر رگرسیونی و کمکی استفاده شده است.

در شکل ۹ نمودار مقادیر پیش‌بینی شده برای  $TDS_1$  تا زمان ۵۰، و در جدول ۴ مقادیر پیش‌بینی شده همراه با بازه‌ی اطمینان آن نشان داده شده است. در ادامه به بررسی و تحلیل و پیش‌بینی مقادیر آینده‌ی متغیرهای معرفی شده می‌پردازیم. مراحل و شیوه‌ی تحلیل همانند روش بالا است. برای جلوگیری از تکرار مطالب نتایج به‌طور خلاصه بیان می‌شوند.

پس از برازش مدل چندجمله‌یی مرتبه‌ی اول و مدل چندجمله‌یی مرتبه‌ی دوم با ضرایب تخفیف مختلف به سری زمانی  $TDS_2$ ، نشانه‌هایی از عدم برازش مدل به داده‌ها مشاهده شد. به طوری که در مدل چندجمله‌یی مرتبه‌ی اول با  $\delta_T = 0.765$

به استثناء چند نقطه‌ی غیرعادی در اوایل سری که باعث چند نقطه با قدرمطلق خطای زیاد شده، مشاهده می‌شود که با گذشت زمان به علت خاصیت پویایی، مانده‌ها کم و کم‌تر شده‌اند. یعنی مدل رگرسیونی به خوبی به داده‌ها برازش یافته است. خاصیت مدل پویا این است که همواره برای برآورد هر نقطه، از اطلاعات قبل از آن نقطه استفاده می‌کند. بدیهی است که در ابتدای کار تعداد مشاهدات قبلی و بنابراین اطلاعات پیشین اندک است و شاهد نقاطی با قدرمطلق خطای زیاد هستیم، ولی با گذشت زمان اطلاعات پیشین بیشتر می‌شود و به نیکویی برازش مدل کمک می‌کند.

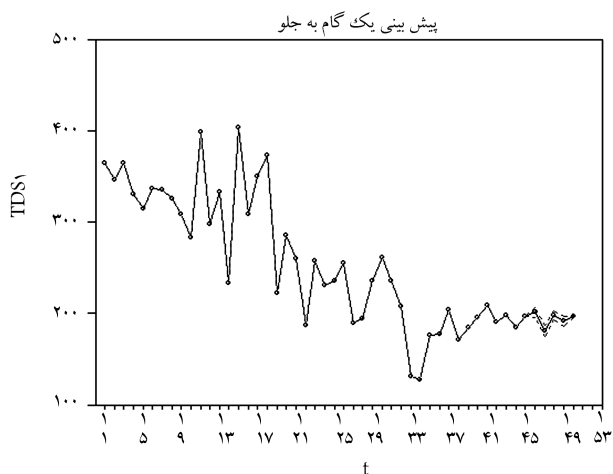
شکل ۸ نمودار پیش‌بینی یک گام به جلو و بازه‌های اطمینان ۹۵٪ با  $\delta_R = 0.795$  و  $\delta_T = 0.797$  را نشان می‌دهد. قرار گرفتن همه‌ی نقاط در داخل بازه‌ی اطمینان، کاهش تدریجی مقدار  $S_e$  و کوتاه‌تر شدن طول بازه‌ی اطمینان مقادیر پیش‌بینی شده، همگی نشان‌دهنده‌ی برازش خوب مدل به داده‌ها است که جدول ۳ الف قابلیت صحت مدل را نشان می‌دهد. چنان که در جدول ۳ ب ملاحظه می‌شود، مقادیر پیش‌بینی شده با مقادیر واقعی اختلاف معناداری ندارند، لذا مدل از لحاظ دقت و صحت مدل خوبی است.



شکل ۸. سری مشاهدات (نقطه‌ها) و سری مقادیر پیش‌بینی یک گام به جلو (خط پر) مدل رگرسیونی از متغیر  $TDS_1$  با بازه‌ی اطمینان ۹۵٪.

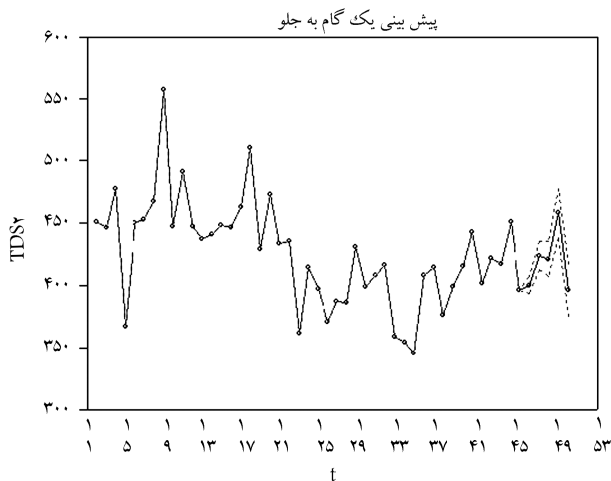
جدول ۳. مقایسه‌ی مقادیر پیش‌بینی شده با مقادیر واقعی متغیر.

ردیف داده‌ها	مقادیر واقعی	مقادیر پیش‌بینی شده	فاصله‌ی اطمینان ۹۵٪
۳۸	۱۷۱٫۶	۱۷۰٫۴	۵٫۳۴
۳۹	۱۸۴٫۳	۱۸۲٫۹	۵٫۸۶
۴۰	۱۹۵٫۸	۱۹۴٫۳	۶٫۲۱
۴۱	۲۰۹	۲۰۷٫۵	۶٫۹۲
۴۲	۱۹۰٫۶	۱۸۹	۷٫۴۸
۴۳	۱۹۷٫۸	۱۹۶٫۲	۸٫۱۲
۴۴	۱۸۵	۱۸۳٫۳	۸٫۹۸
۴۵	۱۹۶٫۵	۱۹۴٫۶	۹٫۳۵

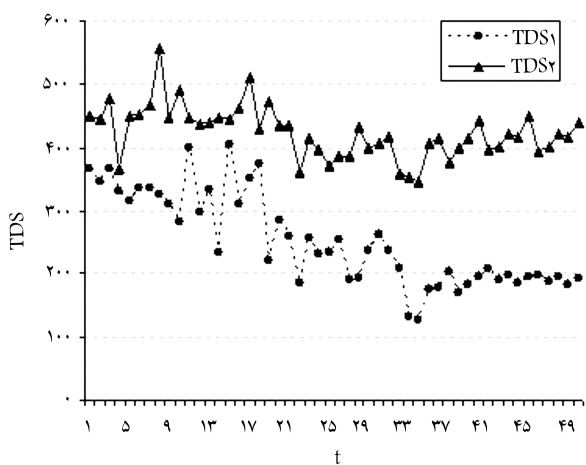


شکل ۹. سری مقادیر پیش‌بینی شده‌ی بر مبنای مدل رگرسیونی.





شکل ۱۰. سری مقادیر پیش‌بینی شده‌ی بر  $TDS2$  مبنای مدل رگرسیونی.



شکل ۱۱. سری‌های  $TDS$  همراه با مقادیر پیش‌بینی شده‌ی آنها در دو ایستگاه مورد مطالعه.

با بررسی داده‌های مشاهده شده، دیده شد که اکسیژن محلول در آب ( $DO$ ) به دمای آب ( $T$ ) وابسته است. مثلاً همبستگی سری زمانی  $DO \setminus$  با سری زمانی  $T \setminus$  برابر با  $r = -0.76$  است. بنابراین پس از برازش مدل چندجمله‌یی مرتبه‌ی اول و مدل چندجمله‌یی مرتبه‌ی دوم به سری زمانی  $DO \setminus$  و مشاهده‌ی نشانه‌هایی از نامناسب بودن مدل، سرانجام مدل رگرسیونی پویا با متغیر پاسخ  $DO \setminus = Y$  و متغیر ورودی  $T \setminus = X$  در نظر گرفته شد. در این حالت ضرایب تخفیف  $\delta_R = 0.9$  و  $\delta_T = 0.95$ ، ضرایب مناسبی هستند که با اعمال آنها، کم‌ترین  $MAD$  برابر  $0.534$  و کم‌ترین  $MAD$  برابر  $0.521$ ،  $(MSE = 0.534, 0.521)$  به دست می‌آید. به کمک نمودار مانده‌ها و سایر نمودارها، نیکویی برازش مدل تأیید می‌شود. در اینجا فقط مقادیر پیش‌بینی شده همراه با بازه‌ی اطمینان آن و نمودار مقادیر پیش‌بینی شده برای  $DO \setminus$  تا زمان ۵۰ آورده شده است، که به ترتیب در جدول ۶ و شکل ۱۲ نشان داده شده است.

همچنین سری زمانی  $DO2$  با ضریب همبستگی  $r = -0.69$  به سری زمانی  $T2$  وابسته است. سرانجام پس از برازش مدل‌های مختلف و بررسی آنها، مدل

جدول ۴. مقادیر پیش‌بینی شده‌ی همراه با بازه‌ی اطمینان ۰/۹۵.

ردیف	مقادیر پیش‌بینی شده	فاصله اطمینان ۹۵٪
۴۶	۲۰۱	۵,۵۷۸
۴۷	۱۸۰,۷	۵,۷۲۹
۴۸	۱۹۸,۱	۶,۴۶۱
۴۹	۱۹۱,۳	۶,۹۵۷
۵۰	۱۹۶,۵	۷,۲۸۷

میزان  $MSE = 1292.2$  و در مدل چندجمله‌یی مرتبه‌ی دوم با  $\delta_T = 0.8$ ، میزان  $MSE = 1722.9$  به دست آمد. در هر یک از مدل‌ها، وجود بسیاری از مجموعه‌ی نقاط با قدر مطلق خطای زیاد در نمودار مانده‌ها، وجود نقاطی در خارج از بازه‌ی اطمینان در نمودار پیش‌بینی یک گام به جلو، افزایش تدریجی مقدار  $S_t$  و طولانی‌تر شدن بازه‌ی اطمینان مقادیر پیش‌بینی شده، همگی حاکی از عدم برازش خوب مدل به داده‌ها بود.

همچنین با توجه به همبستگی زیاد سری زمانی  $TDS2$  با سری زمانی  $T2$ ، مانند حالت قبل مدل رگرسیونی پویا را که در آن  $TDS2 = Y$  متغیر پاسخ، و  $T2 = X$  متغیر کمکی ورودی است به داده‌ها برازش دادیم. براساس نتایج حاصله، با ضرایب تخفیف  $0.98$  و  $0.96$  و  $\delta_R = 0.98$  و  $\delta_T = 0.96$  کم‌ترین مقادیر  $MSE$  و  $MAD$  به ترتیب برابر با  $815.6$  و  $13.27$  است و مشاهده شد که این معیارها در مدل رگرسیونی نسبت به مدل‌های قبل خیلی کاهش یافته‌اند. همه‌ی نمودارها نیز نیکویی برازش مدل را تأیید می‌کنند که از آوردن نمودارها در اینجا صرف‌نظر می‌کنیم. نمودار مقادیر پیش‌بینی شده برای  $TDS2$  تا زمان ۵۰ در شکل ۱۰ نشان داده شده است و مقادیر پیش‌بینی شده به همراه بازه‌ی اطمینان آن در جدول ۵ آمده است. مقادیر آینده‌ی  $T2$  که برای پیش‌بینی  $TDS2$  لازم‌اند، مانند قسمت قبل انتخاب شده‌اند.

برای مقایسه‌ی تفاوت کیفیت آب در بالادست با کیفیت آن در پایین دست، پس از ورود پساب‌های شهری و صنعتی به این رودخانه، سری‌های متغیر  $TDS$  این دو ایستگاه همراه با مقادیر پیش‌بینی شده‌ی آنها در نمودار شکل ۱۱ رسم شده است. مشاهده می‌شود که ورود پساب‌ها باعث افزایش چشم‌گیری در مقادیر  $TDS$  شده که در مقایسه‌ی استاندارد آب آشامیدنی ایران که مطلوب آن  $50$  میلی‌گرم بر لیتر و مجاز آن  $200$  میلی‌گرم بر لیتر است<sup>[۲۰]</sup> مشکلی بروز نمی‌کند. در استاندارد تخلیه‌ی پساب خروجی، مصوب سازمان محیط زیست ایران، برای مصارف کشاورزی حد آستانه برای غلظت  $TDS$  مشخص نشده است.<sup>[۲۱]</sup>

جدول ۵. مقادیر پیش‌بینی شده‌ی  $TDS2$  همراه با بازه‌ی اطمینان ۰/۹۵.

ردیف	مقادیر پیش‌بینی شده	فاصله اطمینان ۹۵٪
۴۶	۴۰۰	۶,۳۲۵
۴۷	۴۲۳,۹	۶,۸۱۲
۴۸	۴۲۱	۷,۰۰۳
۴۹	۴۵۷,۸	۷,۴۰۹
۵۰	۳۹۶,۳	۷,۲۱۸

جدول ۶. مقادیر پیش‌بینی شده‌ی  $DO_1$  همراه با بازه‌ی اطمینان ۹۵٪.

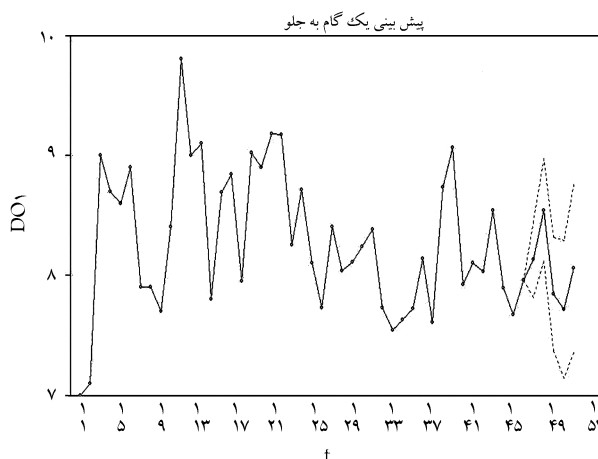
ردیف	مقادیر پیش‌بینی شده	فاصله اطمینان ۹۵٪
۴۶	۸٫۱۳	۰٫۳۱۶
۴۷	۸٫۵۴	۰٫۴۲۶
۴۸	۷٫۸۴	۰٫۴۷۲
۴۹	۷٫۷۱	۰٫۵۷۳
۵۰	۸٫۰۵	۰٫۶۹۸

جدول ۷. مقادیر پیش‌بینی شده همراه با بازه‌ی اطمینان ۹۵٪.

ردیف	مقادیر پیش‌بینی شده	فاصله اطمینان ۹۵٪
۴۶	۵٫۴۲۲	۰٫۸۶۷
۴۷	۵٫۱۰۴	۰٫۸۸۵
۴۸	۵٫۳۴۳	۰٫۸۷۸
۴۹	۵٫۳۹۰	۰٫۸۸۲
۵۰	۵٫۱۸۳	۰٫۸۹۵



شکل ۱۳. سری مقادیر پیش‌بینی شده‌ی  $DO_2$  بر مبنای مدل رگرسیونی.



شکل ۱۲. سری مقادیر پیش‌بینی شده‌ی  $DO_1$  بر مبنای مدل رگرسیونی.

### مدل چندجمله‌یی مرتبه‌ی دوم برای $pH_1$ و $pH_2$

اما برای متغیرهای  $pH_1$  و  $pH_2$ ، پس از ارزیابی مدل چندجمله‌یی درجه‌ی اول و مدل چندجمله‌یی درجه‌ی دوم، مدل چندجمله‌یی درجه‌ی دوم برای هر دو آنها مناسب تشخیص داده شد. مدل چندجمله‌یی درجه‌ی دوم که در بالا توضیح داده شده است دارای دو پارامتر سطح و رشد ( $\beta_t, \mu_t$ ) است. برای انتخاب ضریب تخفیف مناسب بازهم مدل‌های مختلفی به داده‌ها برازش داده شد و از بین مدل‌های برازش داده‌شده مدلی که دارای کم‌ترین  $MSE$  بود انتخاب شد. در اینجا برای اجتناب از تکرار مطالب مراحل ارزیابی مدل‌ها تا رسیدن به بهترین مدل، آورده نمی‌شود.

برای متغیر  $pH_1$  با توجه به ملاک‌های ذکرشده، مدل چندجمله‌یی درجه‌ی دوم با ضریب تخفیف  $0.09$  ( $\delta_T = 0.09$ ) به دست آمد. مقادیر پیش‌بینی شده همراه با بازه‌ی اطمینان ۹۵٪ آن در جدول ۸ آمده است. چنان‌که ملاحظه می‌شود تقریباً همه‌ی نقاط در داخل بازه‌ی اطمینان قرار گرفته‌اند. همچنین برای متغیر  $pH_2$  طبق ملاک‌های بیان‌شده، مدل چندجمله‌یی درجه‌ی دوم با ضریب تخفیف  $0.75$  ( $\delta_T = 0.75$ ) به دست آمد. مقادیر پیش‌بینی شده همراه با بازه‌ی اطمینان ۹۵٪ آن در جدول ۹ آمده است. چنان‌که ملاحظه می‌شود تقریباً تمامی نقاط در داخل بازه‌ی اطمینان قرار گرفته‌اند.

اگرچه ارزیابی و بررسی مناسب بودن مدل‌های برازش داده شده به داده‌ها، با بررسی باقی‌مانده‌ها و اجزای دیگر به‌تنهایی برای اطمینان به درستی و خوب بودن مدل کافی است، برای اطمینان کامل از درستی و اعتبار نتایج حاصل، قابلیت

رگرسیونی پویا با متغیر پاسخ  $DO_2 = Y$  و متغیر ورودی  $T_2 = X$  و ضرایب تخفیف  $\delta_T = 0.099$  و  $\delta_R = 0.099$  که دارای کم‌ترین  $MSE$  و  $MAD$  به‌کمک نمودار مانده‌ها و سایر نمودارها، نیکویی برازش مدل تأیید می‌شود.<sup>[۱۶]</sup> مقادیر پیش‌بینی شده و نمودار آنها به‌ترتیب در جدول ۷ و شکل ۱۳ آمده است.

با توجه به داده‌های اندازه‌گیری شده توسط محققین میزان اکسیژن محلول در ماه‌های دی، بهمن و اسفند به علت کاهش درجه‌ی حرارت و افزایش میزان بارندگی رو به افزایش بوده است، همچنین ممکن است ورود فاضلاب‌ها به شبکه‌ی مورد مطالعه باعث بر این روند اثر بگذارد و باعث کاهش موضعی اکسیژن محلول شود. اما توجه به داده‌ها از ابتدای بهار تا پایان آن نشان می‌دهد که علی‌رغم وجود بارندگی‌های زیاد در بهار، به علت افزایش درجه حرارت آب و کاهش تدریجی بارندگی، روند اکسیژن محلول تقریباً روندی کاهشی است که صحت مشاهدات را تأیید می‌کند. مقایسه‌ی متوسط مقادیر  $DO_1$  و  $DO_2$  (جدول ۱ الف)، نشان‌دهنده‌ی کاهش حدود دو میلی‌گرم بر لیتر اکسیژن محلول بین دو ایستگاه است که بیان‌گر ورود آلاینده‌ها به شبکه است. به نظر می‌رسد با یک مدیریت صحیح شامل تکمیل شبکه‌ی فاضلاب شهری و جلوگیری از ورود آلاینده‌ها به شبکه بتوان از آب با اکسیژن بالاتر که یکی از نشانه‌های کیفیت آب است، در مصارف شهری استفاده کرد. با ایجاد تهیدات مذکور می‌توان با استفاده از مدل یادشده، میزان اکسیژن محلول در آب را در کوتاه‌مدت پیش‌بینی کرد که اثر بخشی آن تهیدات در آن ملاحظه خواهد شد.

تغییرات  $pH$  نسبت به دامنه‌ی تغییرات  $pH$  اندکی افزایش یافته که این تغییرات ممکن است ناشی از ورود فاضلاب‌های خانگی به شبکه باشد.

اما در مورد کاربرد احتمالی مدل‌ها و پیش‌بینی‌های حاصل از این مطالعه می‌توان گفت که چون از این آب عمدتاً برای کشاورزی استفاده می‌شود و نیز با توجه به آلودگی‌های مختلف (انسانی و صنعتی) که وارد شبکه‌ی مورد مطالعه می‌شود، ضروری است برای تصفیه‌ی آب برای مصارف کشاورزی، صنعتی و... پارامترهای کیفی آب دست‌کم برای آینده‌ی کوتاه‌مدت پیش‌بینی شود تا براساس آن طرح تصفیه‌ی مناسبی انتخاب شود. لذا مدل‌های ساخته‌شده در این مطالعه می‌توانند آینده‌ی پارامترهای کیفی آب را در کوتاه‌مدت با اطمینان قابل قبول پیش‌بینی کنند که ممکن است کمک خوبی به طراحی و بهره‌برداری در تصفیه‌ی مناسب باشد. اگر قرار باشد تعدادی آلوده‌کننده‌ی نقطه‌ی کنترل‌شده (به‌طور مثال تصفیه‌خانه‌ی شوش و صاحبقرانیه) وارد شبکه شود، به‌طوری که تغییرات چندانی در کیفیت آب شبکه ایجاد نکنند، پیش‌بینی آینده‌ی پارامترهای کیفی آب در مدل‌های مورد مطالعه ممکن است راهنمایی برای تعیین میزان ورود آلاینده‌های نقطه‌ی کنترل‌شده به شبکه باشد.

### نتیجه‌گیری

با حرکت از ایستگاه بالادست به سمت ایستگاه پایین دست، به علت ورود قسمتی از فاضلاب‌های خانگی به شبکه، بر میزان مواد محلول افزوده می‌شود. مدل مناسب برای مواد محلول هم برای هر دو ایستگاه، مدل رگرسیون پویا است که با اطمینان قابل قبولی برازش داده شده است. ضرایب تخفیف انتخابی برای مواد محلول بالادست عبارت‌اند از:  $0.95$  و  $0.97$ ، و براین اساس  $MSE$  برابر  $716.5$  و  $MAD$  برابر  $12.21$  خواهد بود. مشاهده شد که این معیارها در مدل رگرسیونی نسبت به مدل‌های دیگر مورد بررسی کاهش چشم‌گیر دارد. همه‌ی نمودارها نیز نیکویی برازش مدل را تأیید می‌کنند و مقادیر واقعی با مقادیر پیش‌بینی شده تفاوت معناداری ندارند.

در مورد اکسیژن محلول در هر دو ایستگاه نیز، مدل رگرسیونی پویا با متغیر رگرسیونی دما، برای ایستگاه پایین دست مدل مناسبی است، ضرایب تخفیف  $\delta R = 0.99$  و  $\delta T = 0.99$  ضرایب مناسبی هستند که کم‌ترین  $MSE$  و  $MAD$  از آنها حاصل می‌شود:  $MSE = 0.618$ ،  $MAD = 0.607$ . به کمک نمودار مانده‌ها و سایر نمودارها، نیکویی برازش مدل تأیید می‌شود.

اما برای متغیرهای  $pH$  و  $pH$ ، پس از ارزیابی مدل چندجمله‌ی درجه‌ی اول و مدل چندجمله‌ی درجه‌ی دوم، مدل چندجمله‌ی درجه‌ی دوم برای اسیدیته‌ی ایستگاه بالادست و پایین دست مناسب تشخیص داده شد. برای مقادیر پیش‌بینی شده همراه با بازه‌ی اطمینان  $95\%$ ، تقریباً تمامی نقاط در داخل بازه‌ی اطمینان قرار گرفته‌اند. همچنین تمام معیارها و نمودارها نیز نیکویی برازش مدل را تأیید می‌کنند و مقادیر واقعی با مقادیر پیش‌بینی شده‌ی آنها تفاوت معناداری ندارند.

چون پیش‌بینی کوتاه‌مدت کیفیت آب برای مصارف گوناگون (کشاورزی، شرب و صنعتی) ضرورت بیشتری دارد، مثلاً در فصل آبیاری دانستن مواد محلول آب یک ضرورت است، و چون مدل‌های به‌دست آمده، آینده‌ی کوتاه‌مدت را به خوبی پیش‌بینی می‌کنند، از نتایج آن ممکن است در مدیریت آبیاری کشاورزی استفاده شود. برتری دیگر و مهم مدل‌های پویای پی‌زی ساخته‌شده این است که محدودیتی در استفاده از داده‌هایی که نمونه‌گیری آنها در فاصله‌های زمانی نابرابر صورت گرفته ندارند.

جدول ۸. مقادیر پیش‌بینی شده‌ی  $pH$  همراه با بازه‌ی اطمینان  $95\%$ .

ردیف	مقادیر پیش‌بینی شده	فاصله اطمینان $95\%$
۴۶	۷,۹۶۴	۰,۳۲۸ / ۷,۳۲۰
۴۷	۷,۹۰۹	۰,۳۸۰ / ۷,۱۶۳
۴۸	۷,۸۵۴	۰,۴۴۷ / ۶,۹۷۷
۴۹	۷,۷۹۸	۰,۵۲۷ / ۶,۷۶۵
۵۰	۷,۷۴۳	۰,۶۱۷ / ۶,۵۳۲

جدول ۹. مقادیر پیش‌بینی شده‌ی  $pH$  همراه با بازه‌ی اطمینان  $95\%$ .

ردیف	مقادیر پیش‌بینی شده	فاصله اطمینان $95\%$
۴۶	۸,۰۵۵۴	۰,۲۹۸ / ۷,۴۶۹
۴۷	۸,۰۵۱۹	۰,۳۱۱ / ۷,۴۴۱
۴۸	۸,۰۴۸۴	۰,۳۲۵ / ۷,۴۱۱
۴۹	۸,۰۴۴۹	۰,۳۴۰ / ۷,۳۷۷
۵۰	۸,۰۴۱۴	۰,۳۵۷ / ۷,۳۴۱

جدول ۱۰. مقایسه‌ی مقادیر پیش‌بینی شده با مقادیر واقعی متغیرهای  $pH$  و  $pH$ .

ردیف داده‌ها	مقادیر واقعی $pH$	مقادیر پیش‌بینی شده‌ی $pH$	مقادیر واقعی $pH$	مقادیر پیش‌بینی شده‌ی $pH$
۳۸	۸,۱۷	۸,۰۵	۸,۰۶	۸,۰۸
۳۹	۸,۲۷	۸,۱۳	۸,۳۱	۸,۱۴
۴۰	۸,۲۴	۸,۲۳	۸,۱۱	۸,۱۶
۴۱	۸,۲۴	۸,۲۷	۸,۴۷	۸,۲۶
۴۲	۸,۲۸	۸,۲۹	۸,۲۵	۸,۲۴
۴۳	۸,۱۴	۸,۲۲	۸,۰۸	۸,۱۲
۴۴	۸,۱۲	۸,۱۸	۷,۹۳	۸,۰۵
۴۵	۷,۸۲	۷,۹۶	۷,۸۱	۸,۰۱

صحت مدل‌ها بررسی شده‌اند. فرض شده که تعداد داده‌ها به‌جای ۴۵، مثلاً ۳۷ تا است. سپس برای هر یک از متغیرها، مدل انتخابی بر روی داده‌های ۱ تا ۳۷ برازش داده شده و مقادیر ۳۸ تا ۴۵ پیش‌بینی شده است. سپس مقادیر پیش‌بینی شده‌ی ۳۸ تا ۴۵ با مقادیر واقعی موجود مقایسه می‌شوند. این بررسی در جدول ۱۰ برای  $pH$  و  $pH$  آمده است. ملاحظه می‌شود که تفاوت معناداری بین مقادیر پیش‌بینی شده و مقادیر واقعی وجود ندارد.

با مشاهده‌ی داده‌های  $pH$ ،  $pH$  این شبکه تغییرات چندانی ندارد و بیشتر مقادیر مشاهده شده‌ی آن بین ۷ تا ۸ است. این امر ممکن است نشان‌دهنده‌ی این باشد که آلودگی اسیدی عمده‌ی به شبکه وارد نمی‌شود و نیز آب اندکی خاصیت قلیایی دارد که می‌تواند ناشی از ورود آب‌های زیرزمینی به شبکه باشد. دامنه‌ی

	علائم اختصاری
The Random Variable under Consideration	متغیر تحت بررسی $Y$
The Observation Series at $t$ Time	سری مشاهداتی در زمان $t$
The State of Knowledge at $t$ Time	مجموعه اطلاعات در زمان $t$
The Probability Density Function	تابع چگالی $P(\cdot)$
The Probability Density Function for the Conditional upon $Y$ Variable $D$ Information	تابع چگالی متغیر تصادفی $Y$ به شرط مجموعه اطلاعاتی $D$ $P(Y D)$
to be Normally Distributed $Y$ with Mean $\mu$ and Variance $V$	متغیر $Y$ دارای توزیع نرمال با میانگین $\mu$ و واریانس $V$ $Y \sim N(\mu, V)$
Degree $n$ $t$ Distribution with of Freedom	توزیع $t$ با $n$ درجه آزادی $T_n(m, C)$
Gamma Distribution with $a, b$ Parameters	توزیع گاما با پارامترهای $a$ و $b$ $G(a, b)$
Expectation	امید ریاضی $E[\cdot]$
Variance	واریانس $V[\cdot]$
Covariance	کوارانس $COV[\cdot, \cdot]$
Standard Residual	مانده‌ی استاندارد $u_t$
Autoregressive Integrated Moving Average	مدل میانگین متحرک جمع‌بسته‌ی اتورگرسیو $ARIMA$
Bayesian Analysis of Time Series Software	نرم‌افزار تحلیل بیزی سری‌های زمانی $BATS$
Total Dissolved Solids	کل مواد محلول $TDS$
Electrical Conductivity	هدایت الکتریکی $EC$
Dissolved Oxygen	اکسیژن محلول $DO$
Temperature	درجه حرارت $T$
Acidity	اسیدیته $pH$
Mean Square Error	میانگین توان دوم خطاها $MSE$
Mean Absolute Deviation	میانگین قدرمطلق انحرافات $MAD$

### پانوشت

1. exponential smoothing
2. dynamic linear model (DLM)
3. evolution
4. first order polynomial
5. predictor function
6. linear growth model
7. adaptive coefficient
8. discount factor
9. bayesian analysis of time series Soft Ware (BATS)

### منابع

1. Thomann, R.V. "Time series analysis of water quality data", *J. Sanit. Eng. Div. Proc. ASCE*, pp. 1-23 (1967).
2. Carlson, R.; McCormick, A. and Walts, D. "Application of linear random models to four annual stream flow time series", *Water Resource Res*, **6**, pp. 1070-1078 (1970).
3. McMichale, F. and Hunter, J. "Stochastic modeling of temperature and flow in rivers", *Water Resour. Res.* **8**, pp. 87-98 (1972).
4. Huck, P.M. and Farquhar, G.J. "Water quality models using Box and Jenkins method", *J. Environ. Eng. Div. ASCE*, **100**, pp. 733-753 (1974).
5. Box, G. and Jenkins, G., *Time series analysis, forecasting and control*, Holden -Day. San Francisco, California. U.S.A. (1976).
6. Lohani, B.N. and Wang, M.M. "Water quality data analysis in Chung Kang River", *J. Environ. Eng. Div. ASCE*, **113**, pp. 186-195 (1987).
7. Jayawardena, A. and Lai, F. "Time series analysis of water quality data in Pearl River", *J. Environ. Eng. Div. ASCE*, **113**, pp. 590-606 (1987).
8. Jayawardena, A. and Lai, F. "Water quality forecasting using and adaptive ARIMA modeling approach", *Proc. Int Symp. Env. Hydra, Hong Kong*, pp. 1121-1127 (1991).

9. MacLeod, C. and Whitfield, P.H. "Seasonal and long-term variations in water quality of the columbia river at revelstoke", *B.C. Northwest Science*, **70**(1), pp. 55-65 (1996).
10. Caissie, D.; EL-Jabi, N. and St-Hilaire, A. "Stochastic modeling of water temperature in a small stream using air to water relations", *Can. J. Civ. Eng.* **25**, pp. 250-260 (1998).
11. Hounslow, W., *Water Quality Data, Analysis and Interpretation*, Lewis Publishers, New York (1995).
12. Asadollahfardi, G.R. "Analysis of surface water quality in Tehran", *Water Qual. Res. J. Canada*, **37**(2), pp. 489-511 (2002).
13. Kurunc, A.; Yurekli, K. and Cevik, O. "Performance of two stochastic approaches for forecasting water quality and streamflow data from Yesilirmak River, Turkey", *Environmental Modelling & Software*, **20**, pp. 1195-1200 (2005).
14. Box, G.E.P. and Tiao, G.C., *Bayesian Inference in Statistical Analysis*, Addison Wesley, Massachusetts (1973).
15. West, M. and Harrison, P.J., *Bayesian forecasting and dynamic models*, Springer-Verlag, New York (1989).
16. Samadi, S.Y. "Application of bayesian time series models in the analysis of physical and chemical qualities of surface water in North and South East of Tehran", Master of Science Thesis, Shahid Beheshti Universty (1385).
17. Quintana, J.M. "A dynamic linear matrix-variate regression models", Research report 83, Department of Statistics, University of Warwick (1985).
18. West, M.; Harrison, P.J. and pole, A. "BATS bayesian analysis of time series", *The Professional Statistician*, **6**, pp. 43-46 (1987).
19. pole, A.; West, M. and Harrison, P.J., *Applied Bayesian forecasting and time series analysis*, Chapman and Hall, New York (1994).
20. Drinking Water Quality Standards, Management and Planning Organization Publication (2002).
21. Collection of Environmental Rules and Regulations, Department of Environment of Iran Publication (2000).

