



## بهبود نتایج پیش بینی BOD رودخانه ها بر پایه نوپرزدایی با مقایسه مدل های موجک عصبی، برنامه ریزی ژنتیک، شبکه عصبی و رگرسیون خطی (مطالعه موردی: ایستگاه خروجی سد کرج)

طاہر رجایی<sup>\*</sup>، حمیدہ جعفری، رقیہ رحیمی

گروه مهندسی عمران، دانشگاه قم، قم، ایران

### تاریخچه داوری:

دریافت: ۶ خرداد ۱۳۹۳  
بازنگری: ۱ شهریور ۱۳۹۴  
پذیرش: ۲۲ مرداد ۱۳۹۵  
ارائه آنلاین: ۲۱ اسفند ۱۳۹۵

### کلمات کلیدی:

اکسیژن خواهی بیولوژیکی  
برنامه ریزی ژنتیک  
تبدیل موجک  
رودخانه کرج  
شبکه عصبی  
نوپرزدایی

**چکیده:** در این مطالعه مدل های شبکه عصبی مصنوعی، رگرسیون خطی چند متغیره، برنامه ریزی ژنتیک و ترکیب شبکه عصبی-موجک برای پیش بینی اکسیژن خواهی بیوشیمیایی ماهانه آب (BOD) در ایستگاه خروجی سد کرج بررسی شد و تأثیر پیش پردازش داده ها روی عملکرد مدل ها به وسیله تجزیه موجک مورد تحقیق قرار گرفت. به این منظور در مدل پیشنهادی اول، سری زمانی BOD مشاهداتی بوسیله توابع تبدیل مختلف در سطوح مختلفی به زیر سری ها تجزیه شدند و به عنوان ورودی در مدل شبکه عصبی قرار گرفتند. در مدل پیشنهادی دوم، سری زمانی BOD در ده سطح تجزیه شد. سپس حاصل جمع جبری زیرسری های مؤثر به عنوان ورودی مدل شبکه عصبی برای پیش بینی ماه آینده BOD در نظر گرفته شد. نتایج نشان داد که عملکرد پیش بینی مدل های ترکیبی موجک عصبی نسبت به شبکه عصبی، برنامه ریزی ژنتیک و رگرسیون بهتر است. این مدل ترکیبی برای مقادیر بیشینه نیز پیش بینی قابل قبولی را ارائه داده است. همچنین مدل ترکیبی دوم، میانگین قدمطلق خطا را برای مدل های رگرسیون، برنامه ریزی ژنتیک، شبکه عصبی و مدل ترکیبی اول، به ترتیب از ۱/۸۷، ۰/۹۱، ۰/۶۵ و ۰/۴۶ به ۰/۴۴ کاهش داد و ضریب کارایی را از ۰/۲۳، ۰/۵۳، ۰/۷۳ و ۰/۸۱ به ۰/۸۳ افزایش داد.

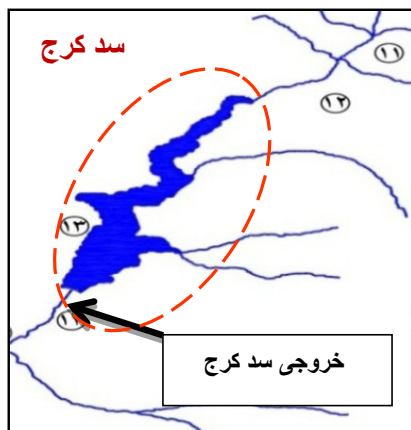
### ۱- مقدمه

کیفیت آب های سطحی یک عامل مؤثر در سلامتی انسان و سیستم های زیست محیطی به خصوص در مناطق شهری می باشد. با توجه به اینکه رودخانه ها با عبور از شهرها بسیاری از آلاینده های رها شده از فاضلاب های خانگی، صنعتی و پساب های کشاورزی را دریافت می کنند، وجود اطلاعات قابل اعتماد در مورد کیفیت آب برای مدیریت آن مؤثر می باشد. شاخص های آلودگی زیست محیطی در یک رودخانه با در نظرگیری شرایط اکسیژن مورد نیاز بیولوژیکی (BOD) قابل ارزیابی است. BOD بیانگر مواد زیستی قابل تجزیه موجود در آب و مقدار اکسیژن لازم برای ثبات بیولوژیکی در آب بوده و جزء اولین پارامترهای مورد نیاز برای ارزیابی کیفیت هر سیستم آبی به شمار می رود. روش های متداول تعیین BOD اغلب روش های بسیار مشکل و توأم با خطاهای اندازه گیری می باشد. از آنجاییکه حفاظت کیفی آب رودخانه ها، نیاز به سرمایه گذاری های اضافی جهت تصفیه پساب ها یا سیستم های جمع آوری و کنترل زه آب ها داشته و از طرفی ممکن است منجر به محدود کردن توسعه فعالیت ها در حوزه رودخانه گردد، لذا می تواند اثرات اقتصادی قابل توجهی داشته باشد. بنابراین وجود یک ابزار به منظور پیش بینی

و نشان دادن شرایط موجود و آتی و نیز محدودیت های لازم جهت دست یابی به استانداردهای کیفی، ضروری به نظر می رسد. برخی از مدل سازی ها و پیش بینی های پارامترهای کیفی بر پایه سری های زمانی پارامترها صورت می گیرد. برخلاف روش های سنتی که با فرض رابطه ای خطی بین مقادیر واقعی و پیش بینی شده عمل می کنند، روش های جدید هوش مصنوعی قادر به برازش غیرخطی داده ها می باشند. امروزه این مدل ها در طیف وسیعی از تحقیقات استفاده می شوند. در این راستا نجاح و همکاران (۲۰۰۹) [۱۰] توانایی شبکه های عصبی را در پیش بینی شاخص های کیفی آب رودخانه جوهر مالزی و توانایی برآورد مقدار EC و TDS و کدورت را مورد بررسی قرار دادند. رجایی و میرباقری [۲] با استفاده از شبکه های عصبی توانستند غلظت بار معلق در زمان آینده را پیش بینی نمایند. کیم و سو [۹] (۲۰۱۵) با استفاده از شبکه های عصبی مصنوعی و روش خوشه بندی کیفیت آب رودخانه را پیش بینی کردند. مکارین اسکای و همکاران (۲۰۱۵) از ترکیب شبکه های عصبی مصنوعی و مدل های عددی معین برای تخمین غلظت رسوب استفاده کردند. اغلب سری های زمانی هیدرولوژیکی به علت تابعیت از تغییرات فصلی، ماهانه یا روزانه و اقلیم منطقه مورد مطالعه، دارای مشخصات غیرایستا، پرش، روند و تغییرات ناگهانی در سری زمانی خود هستند، به همین دلیل برای

<sup>\*</sup>نویسنده عهده دار مکاتبات Trajaee@qom.ac.ir

دقیقه عرض شمالی و ۵۱° و ۶ دقیقه تا ۵۱° و ۳۳ دقیقه طول شرقی قرار دارد. حوضه آبریز رودخانه کرج از شمال به حوضه آبریز رودخانه چالوس، هراز، طالقانرود و از شرق به حوضه آبریز کردان محدود می‌شود. رودخانه کرج از شاخه‌های ولایترو (گاجره)، وارنگرود، سیرا، شهرستانک، مورود، ارنگه، آزادبر، کندوان، کیل، آسارا، سپهسالار و نشترود تشکیل شده که همگی از ارتفاعات البرز سرچشمه می‌گیرند. طول رودخانه حدود ۲۴۵ کیلومتر، مساحت حوضه آبریز آن حدود ۵۰۰۰ کیلومتر مربع، پهنای آن (در طول مسیر متفاوت بوده) مابین ۸ تا ۱۵ متر و عمق آن از ۱ تا ۳ متر متغیر می‌باشد. شیب شاخه اصلی و شاخه‌های فرعی رودخانه بسیار زیاد بوده و به دلیل اینکه در تمام مسیر خود در شیب تند دره‌ها از مناطق سنگلاخی و صخره‌های عبور می‌کند، از نظر ویژگی‌های زیستی و ساختاری بسیار مناسب برای آب شیرین می‌باشد. رودخانه کرج یکی از مهمترین منابع تأمین کننده آب مورد نیاز شهر تهران بوده و به دلیل مناسب بودن آب و هوای منطقه، محل مناسبی برای انجام فعالیت‌های تفریحی محسوب می‌شود که کنترل کیفیت آب آن از اهمیت بسزایی برخوردار است (شکل ۱) [۱].



شکل ۱: موقعیت شماتیک ایستگاه خروجی سد تنظیمی کرج

Fig.1. Schematic position of Karaj regulatory dam outlet station

برای انجام پژوهش حاضر از داده‌های شرکت آب منطقه‌ای استان تهران در محل خروجی سد تنظیمی کرج استفاده شده است. از دوره ی ۱۴۵ ماهه مورد استفاده (تیر ۷۸ تا مرداد ۹۰)، معادل ۸۰٪ داده‌ها برای آموزش و داده‌های ۳۰ ماهه باقی مانده معادل ۲۰٪ داده‌ها به منظور اعتبارسنجی و آزمون مدل‌ها مورد استفاده قرار گرفته است. شکل ۲ سری زمانی BOD در طی دوره آماری را نشان می‌دهد.

تشخیص اطلاعات مخفی سیگنال و بهینه‌سازی وزن‌ها در آموزش شبکه عصبی داده‌های ورودی مدل‌های شبکه عصبی با استفاده از آنالیز موجک ۲ پیش‌پردازش می‌شوند. نورانی و همکاران (۲۰۰۹) [۱۱] با ترکیب شبکه‌های عصبی و آنالیز موجک به مدل‌سازی بارش حوضه ليقوان چای پرداختند. رجایی (۲۰۱۱) [۱۳]، مدل جدید شبکه عصبی موجک را برای بار رسوب معلق روزانه در رودخانه‌ها ارائه داد. به منظور ارزیابی دقت مدل، مدل پیشنهادی با نتایج مدل‌های شبکه عصبی، رگرسیون چندخطی (MLR) و منحنی سنج رسوب (SRC) مورد مقایسه قرار گرفت. مقایسه نشان داد که مدل ترکیبی موجک دقت بالاتری دارد. در تحقیقی دیگر رجایی و همکاران (۲۰۱۱) [۱۴] با مدل ترکیبی شبکه عصبی موجک بار معلق روزانه رودخانه را پیش‌بینی کردند که مدل ترکیبی دقت بالاتری نسبت به مدل عصبی ارائه داد. سینگ (۲۰۱۲) [۱۵] مدل ترکیبی شبکه عصبی موجک را برای پیش‌بینی سیلاب به کار برد که نتایج نشان‌دهنده توانایی بالای مدل ترکیبی نسبت به روش‌های آماری بود. لانگ کین و همکاران (۲۰۱۳) [۸] مدلی برای پیش‌بینی کیفی آب بر اساس موجک و شبکه عصبی ارائه دادند. نتایج نشان داد که استفاده از موجک، دقت مدل را نسبت به مدل شبکه عصبی بالا برده است. نورانی و همکاران (۲۰۱۳) [۱۲] از داده‌های ماهواره‌ای بارش روزانه چهار بارانسج برای پیش‌بینی رواناب روز بعد استفاده کردند. نتایج تحقیق حاکی از برتری مدل ترکیبی موجک بود. شانگ‌این و همکاران (۲۰۱۴) [۱۶] از مدل ترکیبی CPSOLSSVR و الگوریتم موجک برای پیش‌بینی اکسیژن محلول ۳ (DO) رودخانه خرچنگ استفاده کردند که نتایج نشان دهنده بهبود دقت بود. علیزاده و کویانپور (۲۰۱۵) [۴] از مدل ترکیبی شبکه عصبی موجک برای پیش‌بینی کیفیت آب خلیج هیلو در اقیانوس آرام استفاده کردند. نتایج نشان داد که این مدل از دقت بالایی برخوردار است.

تحقیق حاضر از ترکیب شبکه عصبی خودبازگشتی غیرخطی و موجک برای پیش‌بینی سری زمانی پارامتر کیفی BOD رودخانه استفاده می‌کند. برای این منظور، پس از تحلیل نتایج خودهمبستگی بین این پارامتر و تأخیرهای زمانی، ترکیب‌های ورودی مختلفی ایجاد می‌شود و توسط دو مدل پیشنهادی عصبی موجک مدلسازی صورت می‌گیرد. در مدل پیشنهادی اول از چهار نوع تابع موجک در پنج سطح استفاده می‌شود و در مدل دوم با استفاده از نتایج بهترین تابع موجک و انتخاب زیرسری‌های بهینه و جمع جبری آن‌ها، پیش‌بینی انجام می‌شود. نتایج این دو مدل با مدل‌های شبکه عصبی، برنامه‌ریزی ژنتیک و رگرسیون خطی مقایسه می‌شود.

## ۲- مواد و روش‌ها

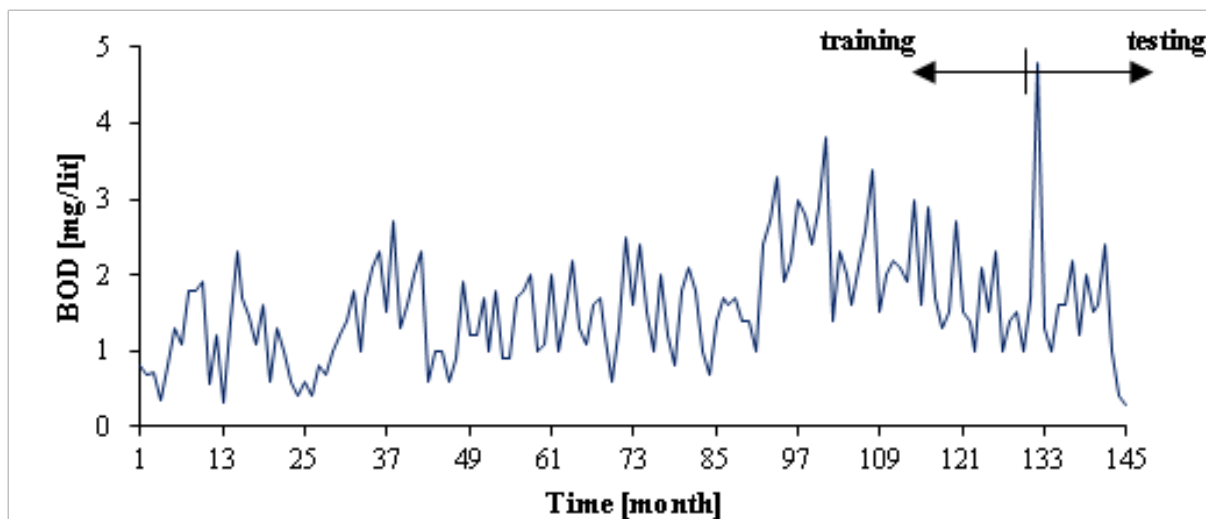
### ۲-۱ منطقه مورد مطالعه

دشت تهران با موقعیت جغرافیایی بین ۳۵° و ۲۸ دقیقه تا ۳۵° و ۴۹

<sup>1</sup> Train

<sup>2</sup> Wavelet analysis

<sup>3</sup> Dissolved oxygen



شکل ۲: سری زمانی BOD ماهانه ایستگاه تنظیمی سد کرج

Fig.2. Monthly BOD time series of Karaj regulatory dam outlet station

## ۲-۲- تحلیل آماری داده‌ها

هر سری زمانی ویژگی‌های رفتاری نهفته‌ای مانند خودهمبستگی، خودشباهتی، خودنسبتی و روند در بازمی تغییرات مقیاسی خود دارد که به این وسیله می‌توان بر تاریخچه‌ی فرآیند آن اشراف پیدا کرده و مقدمات پیش‌بینی رفتار آن را در آینده فراهم ساخت. در جدول ۱ تحلیل آماری داده‌های BOD برای کل دوره مورد مطالعه و نیز به تفکیک، برای داده‌های دوره آموزش و آزمون ارائه شده است. آنالیز ارائه شده شامل میانگین، کمینه، بیشینه، انحراف معیار (Sd)، ضریب چولگی (Cs)، ضریب همبستگی ۱ ماهه (R<sub>1</sub>)، ضریب همبستگی ۳ ماهه (R<sub>3</sub>)، ضریب همبستگی ۴ ماهه (R<sub>4</sub>) و ضریب همبستگی ۶ ماهه (R<sub>6</sub>) می‌باشد.

### جدول ۱: تحلیل آماری سری زمانی داده‌های BOD

Table 1. statistical analysis of BOD data time series

پارامترهای آماری	همه داده‌ها	مجموعه آزمون	مجموعه آموزش
میانگین	۱/۵۷	۱/۵	۱/۵۸
انحراف معیار	۰/۷۳	۰/۷۴	۰/۷۳
ضریب چولگی	۰/۵۵	۰/۷۶	۰/۳۴
کمینه	۰/۳	۰/۳	۰/۳۲
بیشینه	۴/۸	۴/۸	۳/۸
R <sub>1</sub>	۰/۷	۰/۶۸	۰/۷۱
R <sub>2</sub>	۰/۵۴	۰/۵۱	۰/۵۴
R <sub>3</sub>	۰/۵۱	۰/۵۶	۰/۴۸
R <sub>4</sub>	۰/۶۸	۰/۶۶	۰/۶۵

لازم است مشخص گردد پارامتر BOD تا چند تأخیر زمانی به مقدار آن در زمان فعلی وابسته است. با توجه به رفتار و حافظه بلندمدت یا کوتاه‌مدت

سری زمانی، نوع مدل خود همبسته مشخص می‌شود. به این منظور ضرایب خودهمبستگی بین سری زمانی اکسیژن خواهی شیمیایی (BOD<sub>t</sub>) با تأخیرات آن مطابق جدول ۱ محاسبه شد. رابطه ۱ برای محاسبه ضرایب خودهمبستگی بکار گرفته شده که هر چه ضریب همبستگی به عدد ۱ نزدیکتر باشد، نشان‌دهنده سازگاری مناسب بین داده‌هاست.

$$R^2 = \frac{\sum (B_{i(t+1)} - B_{(t+1)})(B_{i(t)} - B_{i(t)})}{\sqrt{\sum (B_{i(t+1)} - B_{(t+1)})^2 \sum (B_{i(t)} - B_{i(t)})^2}} \quad (1)$$

بررسی ضرایب همبستگی نشان می‌دهد که همبستگی پارامتر BOD به ازای تأخیرات ۱ و ۶ ماهه نسبتاً بالا می‌باشد (در جدول ۱ داده‌های اعتبارسنجی جزء داده‌های آموزش لحاظ شده است). بررسی جدول ۱ نشان می‌دهد که شباهت خصوصیات آماری داده‌های دوره آموزش و آزمون، در حد قابل قبولی می‌باشد. ضریب چولگی سری زمانی BOD پایین بوده و برای مدل‌سازی مناسب است [۱۱]. در صورتی که داده‌ها به صورت خام وارد شبکه شوند، به علت تغییرات زیاد داده‌ها تأثیر متفاوتی بر روی شبکه می‌گذارند، لذا لازم است متغیرهای ورودی به بازهای مناسب، منطبق با شکل و رفتار تابع محرک سیگموئید (که مورد استفاده در شبکه عصبی می‌باشد) منتقل شوند. در این تحقیق از رابطه (۲) برای نرمال‌سازی داده‌ها در محدوده [۰، ۱] استفاده شده است [۱۵].

$$X_{norm} = \frac{X_i - X_{min}}{X_{max} - X_{min}} \quad (2)$$

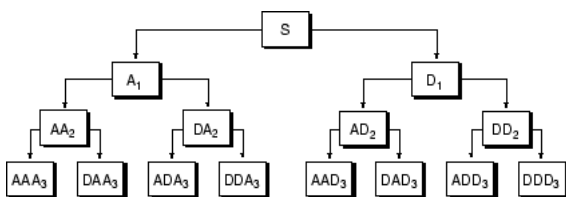
## ۲-۳- شبکه عصبی NAR<sup>۱</sup>

در مدل شبکه عصبی NAR یا خودبازگشتی غیرخطی، تنها داده‌های

<sup>1</sup> Nonlinear Auto Regressive

## ۲-۵- آنالیز موجک

در تبدیل موجک، سیگنال از یک سری از فیلترهای بالاگذر و پایینگذر عبور داده می‌شود. بخش حاصل از عبور سیگنال از فیلتر بالاگذر که شامل اطلاعات فرکانس بالا از جمله نویز می‌باشد، جزئیات نام دارد. بخش حاصل از عبور سیگنال از فیلتر پایینگذر که شامل اطلاعات فرکانس پایین و دربرگیرنده مشخصات هویتی در سیگنال است، تقریباً نامیده می‌شود. فرآیند تجزیه سیگنال به چندین جزء با رزولوشن پائین شکسته می‌شود [۵]. این حالت درخت تجزیه موجک<sup>۱</sup> نامیده می‌شود که در شکل ۴ نشان داده شده است.



شکل ۴: درخت تجزیه سری زمانی

Fig.4. Decomposition tree of time series

## ۲-۶- رگرسیون خطی چندمتغیره<sup>۲</sup> (MLR)

رگرسیون خطی چندمتغیره رابط‌های خطی بین یک متغیر وابسته و یک یا چند متغیر مستقل می‌باشد. شکل معمول معادله MLR به صورت رابطه ۴ می‌باشد که عرض از مبدأ، ضریب رگرسیون متغیرهای مستقل و مقدار پیش‌بینی شده است [۱۰].

$$y = a_0 + \sum a_i x_i \quad (2)$$

## ۲-۷- مدل سازی

با در نظر گرفتن نتایج تحلیل آماری ارائه شده در جداول ۱ و ۲، ترکیب‌های مختلف ورودی برای پیش‌بینی میزان BOD ماه آینده در زمان  $t+1$  به صورت زیر در نظر گرفته شد:

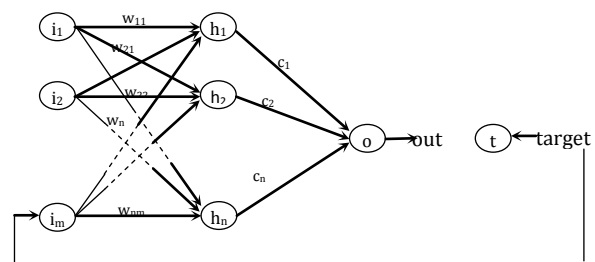
- ۱)  $BDDW_t$
- ۲)  $BDDW_t, BDDW_{t-3}$
- ۳)  $BDDW_t, BDDW_{t-6}$
- ۴)  $BDDW_t, BDDW_{t-3}, BDDW_{t-6}$
- ۵)  $BDDW_t, BDDW_{t-3}, BDDW_{t-6}, BDDW_{t-9}$

تعداد مدل‌ها در روش رگرسیون خطی و برنامه‌ریزی ژنتیک مطابق الگوی فوق ده آنالیز می‌باشد. به منظور مدل‌سازی شبکه عصبی نیز از ترکیب‌های ورودی فوق استفاده می‌شود. تعداد ۷۵ آنالیز برای هر یک از ترکیب‌های ورودی فوق به ازای تعداد نرون‌های لایه پنهان ۱ تا ۱۵. در این تحقیق به منظور آموزش و آزمون از مدل شبکه عصبی سه لایه با روش

یک سری زمانی مورد استفاده قرار می‌گیرد و مقادیر آینده سری زمانی به صورت تابعی از مقادیر گذشته همان سری زمانی پیش‌بینی می‌شوند. به عبارت دیگر:

$$y_t = f(y_{t-1}, y_{t-2}, \dots, y_{t-d}) \quad (3)$$

در این روابط  $d$  تعداد تأخیر است. در هنگام آموزش شبکه که خروجی صحیح در دسترس است، می‌توان معماری حلقه پس‌خور باز (شکل ۳) را به کار برد و خروجی پس‌خور تضمینی استفاده کرد [۶]. ورودی‌های شبکه عصبی به سه دسته داده‌های آموزش، اعتبارسنجی و آزمون تقسیم می‌شوند. روند آموزش شبکه تا زمانی که خطای مربع متوسط (MSE) در مراحل آموزش، اعتبارسنجی و آزمون حداقل گردد، ادامه می‌یابد.



شکل ۳: معماری شبکه عصبی در حالت حلقه پس‌خور باز

Fig.3. Architecture of neural network in state of open feed forward loop

## ۲-۴- برنامه‌ریزی ژنتیک (GP)

مبنای روش شبیه‌سازی GP بر اساس نظریه تکاملی داروین استوار بوده و یکی از شاخه‌های الگوریتم ژنتیک می‌باشد. فرآیند شبیه‌سازی در GP به این صورت است که ابتدا مجموعه‌هایی که می‌توانند جهت انتخاب متغیرها و عملگرها در فرآیند جستجو به کار روند، معرفی می‌شوند. سپس کروموزوم‌ها به وسیله انتخاب یک مجموعه جواب اولیه تصادفی از مجموعه‌های اتصال و توابع ایجاد می‌شوند. سپس تابع هدف متناظر با هر کروموزوم محاسبه می‌شود. سپس مجموع تفاوت مقادیر محاسباتی و مشاهداتی کمینه می‌شود. در مرحله بعد عملگرهای ژنتیک (تزیوج و جهش) اعمال می‌شوند. در ادامه، فرآیند توسعه‌ای تکراری در مورد فرزندان انجام می‌شود تا در نهایت پس از تعداد تکرار مشخصی که روند تغییرات تابع هدف در آن تقریباً ثابت گردید، مقدار بهینه یا نزدیک به بهینه به‌عنوان جواب مسئله استخراج گردد. در این مقاله، برای پیش‌بینی به روش GP از نرم‌افزار Discipulus استفاده شده است. برای معیار توقف در این نرم‌افزار، میزان حداکثر تعداد اجرا و میزان تولید بدون بهبود ۳۰۰ در نظر گرفته شده است. در این نرم‌افزار از ۸۰٪ داده‌ها برای آموزش و ۲۰٪ آن برای صحت‌سنجی و آزمون استفاده می‌شود.

<sup>1</sup> Packet decomposition tree

<sup>2</sup> Multi Linear Regression

داده‌های ورودی بهینه می‌گردد. زیرسری‌هایی که همبستگی بالایی دارند انتخاب شده و مجموع آن‌ها به عنوان سری زمانی جدید در نظر گرفته می‌شود. به این منظور، ضرایب همبستگی بین زیرسری‌های به دست آمده با تأخیر گام‌های زمانی یک، سه، چهار و شش ماهه، مطابق با انتخاب ورودی‌ها در سری زمانی اصلی BOD که در جدول ۱ ارزیابی شده، نیز محاسبه می‌شود.

### ۲-۸- معیارهای ارزیابی مدل

در این تحقیق به منظور ارزیابی عملکرد مدل‌های شبکه عصبی سه شاخص معیاری متفاوت شامل ریشه میانگین مربعات خطای بین داده‌های پیش‌بینی و مشاهده شده<sup>۴</sup> (RMSE)، میانگین قدر مطلق خطا<sup>۵</sup> (MAE)، ضریب کارائی نش‌ساتکلیف یا ضریب تبیین<sup>۶</sup> (NS) به عنوان ارزیابی توان پیش‌بینی مدل، روابط (۸ تا ۱۰) مورد استفاده قرار گرفته است. در این روابط،  $X_i$  و  $Y_i$  به ترتیب آمین داده پارامتر برآوردی و اندازه گیری  $X$  و  $Y$  به ترتیب متوسط داده‌های  $X_i$ ،  $Y_i$  و  $n$  تعداد نمونه‌های مورد ارزیابی است [۱۵].

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (X_i - Y_i)^2}{n}} \quad (8)$$

$$MAE = \frac{\sum_{i=1}^n |X_i - Y_i|}{n} \quad (9)$$

$$E = 1 - \frac{\sum_{i=1}^n (Y_i - X_i)^2}{\sum_{i=1}^n (Y_i - \bar{Y}_i)^2} \quad (10)$$

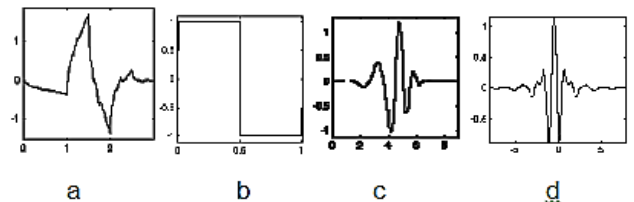
### ۳- نتایج و بحث

در این تحقیق مدل‌های رگرسیون خطی چند متغیره، برنامه‌ریزی ژنتیک و شبکه عصبی با استفاده از ترکیبات ورودی ذکر شده تشکیل شد. سپس بهترین ترکیب با ضریب کارائی بالا و خطای کمتر، توسط مدل پیشنهادی اول شبکه عصبی موجک شبیه‌سازی شد. در اجرای مدل رگرسیون خطی چند متغیره، پارامتر BOD در ماه فعلی به عنوان متغیر وابسته و سایر پارامترها در ترکیبات ورودی به عنوان متغیر مستقل در نظر گرفته شدند. نتایج مربوط به مدل رگرسیون و برنامه‌ریزی ژنتیک در جدول ۲ نمایش داده شده است. آنالیز مدل رگرسیون در نرم‌افزار آماری SPSS انجام گرفت. بهترین نتیجه برای این مدل مربوط به ترکیب ورودی یک با ضریب تبیین ۰/۲۳ می‌باشد. در مدل برنامه‌ریزی ژنتیک بهترین نتیجه برای ترکیب ورودی سوم بدست آمد.

پیش‌خور پس انتشار خطا<sup>۱</sup> (FFBP) با الگوریتم آموزشی لونبرگ-مارکوارت استفاده شد. بر اساس تحقیقات موجود، الگوریتم آموزشی لونبرگ-مارکوارت، به دلیل سرعت، دقت و اطمینان بالای آن در پیش‌بینی‌های هیدرولوژیکی مناسب می‌باشد [۳]. در ساختار شبکه عصبی برای لایه‌های مخفی و خروجی به ترتیب از توابع انتقال سیگموئید<sup>۲</sup> و خطی<sup>۳</sup> استفاده شده و تعداد نورون‌های لایه مخفی به روش سعی و خطا انتخاب گردید.

### ۲-۷-۱- مدل پیشنهادی شبکه عصبی موجک (WANN)

در مدل ترکیبی شبکه عصبی-موجک، زیرسری‌هایی که توسط تبدیل‌های موجک تولید شده‌اند، به عنوان ورودی شبکه عصبی استفاده می‌شوند. استفاده از تبدیل موجک روی داده‌های سری زمانی قبل از ورود به شبکه این امکان را فراهم می‌کند که بتوان از یک آنالیز که تأثیرات کوتاه‌مدت و بلندمدت را در بر می‌گیرد بهره‌مند شد. باید توجه داشت که استفاده از تبدیل‌های موجک و تجزیه سری‌های زمانی الزاماً دقت شبکه عصبی را افزایش نخواهد داد، نکته مهم در انتخاب موجک مادر است. در این تحقیق دو مدل جدید عصبی-موجک پیشنهاد می‌شود. در مدل پیشنهادی اول برای تجزیه سری زمانی داده‌ها به منظور کاستن حجم محاسبات و آنالیزها، بهترین ترکیب ورودی که عملکرد بهتری در مدل‌سازی شبکه عصبی را دارد، برای آنالیز مدل پیشنهادی اول شبکه عصبی-موجک انتخاب می‌شود. به این منظور از چهار نوع موجک که برای هر موجک پنج سطح در نظر گرفته می‌شود و در هر سطح تعداد نورون‌های ۱ تا ۱۵ برای لایه پنهان شبکه عصبی آنالیز می‌شود، استفاده می‌گردد (تعداد ۳۰۰ آنالیز در این روش انجام می‌گیرد). شکل ۵ موجک‌های مورد استفاده در این تحقیق را نشان می‌دهد.



شکل ۵: موجک‌های (a) دابیچز ۲ (b) هار (c) دابیچز ۵ (d) میر  
Fig.5. Wavelets of a) daubechies 2, b) haar, c) daubechies 5, d) Dmey

در مدل پیشنهادی دوم، از موجکی که در روش پیشنهادی اول بهترین نتیجه را دارد برای این روش انتخاب می‌شود. پس از تجزیه این موجک به زیرسری‌ها در ده سطح، ضرایب همبستگی بین زیرسری‌های تولید شده محاسبه می‌شود. با به کار بردن روش همبستگی سری‌های زمانی می‌توان تعداد و تأخیر زمانی مؤثر عناصر ورودی در شبکه‌های عصبی را تعیین کرد، بطوریکه با این روش، زیرسری‌های اضافی و غیرمؤثر حذف و مجموعه

<sup>4</sup> Root Mean Squared Error

<sup>5</sup> Mean Absolute Error

<sup>6</sup> Nash-Sutcliffe model efficiency coefficient

<sup>1</sup> Feed-Forward Back propagation

<sup>2</sup> Logsig

<sup>3</sup> Purelin



جدول ۲: نتایج مدل رگرسیون خطی و برنامه ریزی ژنتیک

Table 2. Results of linear regression model and genetic programming

GP (BOD) (تعداد اجرا = ۱۰۰)					MLR (BOD)					انواع مدل‌ها
۵	۴	*۳	۲	۱	۵	۴	۳	۲	*۱	ترکیب‌ها
۰/۱۹	۰/۲۵	۰/۵۳	۰/۳۲	۰/۴۸	۰/۰۴	۰/۰۹	۰/۱۶	۰/۱۱	۰/۲۳	E
۱/۹۲	۱/۸۶	۰/۹۱	۱/۳۷	۱/۱۴	-	-	۱/۹۶	۲/۰۹	۱/۸۷	MAE (mg/l)
۲/۱۱	۱/۶۲	۱/۲۳	۱/۷۶	۱/۶۶	-	-	۲/۳۳	۲/۶۷	۲/۱۳	RMSE (mg/l)

با معماری شبکه ۱-۹-۲ (یعنی شبکه سه لایه با دو نرون در لایه ورودی، ۹ نرون در لایه مخفی و ۱ نرون در لایه خروجی) و الگوریتم آموزش لوبنبرگ مارکوآرت حاصل شد. همان‌گونه که از جدول ۳ مشخص است، مدل شبکه عصبی نتایج با دقت بالاتری نسبت به مدل رگرسیون خطی دارد.

مدل‌های شبکه عصبی بر اساس تعداد نرون‌های لایه مخفی آموزش داده شد و مورد آزمون قرار گرفت. تعداد نرون‌های لایه مخفی هر شبکه با استفاده از داده‌های موجود و روش سعی و خطا بهینه شد. در اجرای مدل شبکه عصبی NAR، مطابق جدول ۳ بهترین نتیجه در ترکیب ۳

جدول ۳: نتایج مدل شبکه عصبی

Table 3. Results of neural network

ANN (BOD)					انواع مدل‌ها
۵	۴	*۳	۲	۱	ترکیب‌ها
۷	۱۱	۹	۴	۶	تعداد نرون‌های لایه پنهان
۰/۴۸	۰/۵۴	۰/۷۳	۰/۶۸	۰/۷۱	E
۱/۱۲	۰/۸۹	۰/۶۵	۰/۷۵	۰/۶۹	MAE (mg/l)
۱/۶۵	۱/۲۱	۰/۸	۱/۱۵	۰/۸۱	RMSE (mg/l)

پارامتر در آب تغییر می‌کند که این تغییرات در یک حد مشخص می‌باشد. کاهش یا افزایش دما می‌تواند به صورت مستقیم، سطح فعل و انفعالات شیمیایی را تغییر دهد. با کمی دقت در نمودار ۲ نیز مشخص است که آلوده‌ترین زمان رودخانه کرج از نظر افزایش غلظت BOD متعلق به فصل تابستان می‌باشد، چرا که دبی و سرعت جریان رودخانه بسیار پایین بوده و رودخانه امکان خودپالایی را نداشته است. در شرایط غیرعادی از قبیل تخلیه مواد آلاینده به رودخانه یا عوامل اقلیمی مانند بارندگی شدید و سیلاب، میزان تغییرات این پارامتر شدید خواهد بود. این رفتار در حوزه آبریز سد کرج، با احداث مراکز پذیرایی فاقد تصفیه فاضلاب (بعنوان نمونه مساحت‌هایی از هتلهای گچسر و ... که در حریم رودخانه احداث شده‌اند)، ورود فاضلاب سیاه سکونت‌گاه‌های روستایی به چاه‌های جاذب و ورود فاضلاب خاکستری آن‌ها به رودخانه، ورود آلاینده‌های غیر نقطه‌ای متعدد و با منبع نامشخص به درون رودخانه‌ها، ورود پساب‌های کشاورزان به درون رودخانه‌ها بدون انجام تصفیه مناسب، ورود فاضلاب گرمابه‌های فعال به رودخانه که اکثراً در پایین‌دست رودخانه واقع شده‌اند، نمود بیشتری پیدا می‌کند و در سیگنال اصلی این پارامتر نویز ایجاد می‌کند. الگوریتم موجک با شناسایی ورودی‌های تأثیرگذار در سیگنال BOD در ۳ سطح، به نحوی که وزن ورودی‌های کم‌اثر پایین بوده و ورودی‌های مؤثر به صورت برجسته در زیرسری‌های  $d_1$ ،  $d_2$  و  $d_3$  قابل تجزیه باشند، آن را نویززدایی کرده است. این سه زیرسری می‌تواند تأثیر هر یک از عوامل فوق که در افزایش غلظت BOD و در نتیجه آلوده‌تر کردن رودخانه کرج نقش داشته‌اند، باشد.

### ۳-۱- مدل پیشنهادی اول

با توجه به اینکه بیشترین دقت و کمترین خطا بین دو مدل شبکه عصبی و رگرسیون برای ترکیب ورودی سوم مناسب‌ترین عملکرد را ارائه داد لذا پیش‌بینی با مدل ترکیبی عصبی موجک با ترکیب ورودی BODt و ۶-BODt انجام گرفت. به نظر می‌رسد با افزایش سطح تجزیه و ظاهر شدن نمونه‌های فرکانس پایین، عملکرد شبکه عصبی موجک مناسب‌تر باشد ولی مطابق جدول ۴ بهترین نتیجه در سطح تجزیه ۳ با موجک دایبچز ۲ بدست آمده است. در مورد تعداد سطوح می‌توان اینطور استدلال کرد که از آنجایی که تعداد کل داده‌های سری زمانی ۱۴۵ عدد است، بعد از تجزیه در ۵ سطح در هر زیر سری حدود ۵ نمونه وجود خواهد داشت. اگر سطح تجزیه موجک از ۵ فراتر رود، با وجود اینکه نمونه‌های فرکانس پایین که اکثراً هم به سیگنال تعلق دارند ظاهر می‌شوند، تعداد نمونه‌ها در هر زیر سری کمتر شده و در نتیجه با این نمونه‌ها دقت تخمین واریانس نویز کاهش یافته و نتیجه معکوس روی داده‌های شبیه‌سازی شده دارد. افزایش حجم پردازش نیز معضل دیگر آن است. جدول ۴ برتری مدل ترکیبی شبکه عصبی موجک (مدل پیشنهادی اول) را نسبت به مدل‌های شبکه عصبی، برنامه‌ریزی ژنتیک و رگرسیون خطی نشان می‌دهد. غلظت BOD در رودخانه کرج به دلایل مختلف در طول زمان تغییر می‌کند. در شرایط عادی در طول ماه‌های مختلف از سال دمای هوا و در نتیجه دمای آب تغییر می‌کنند. هم‌چنین با توجه به ماه‌های خشک و تر سال و میزان دبی، سرعت آب، شیب تند حوضه رودخانه کرج و شرایط آب و هوایی آن و فرسایش بالای این رودخانه، غلظت این

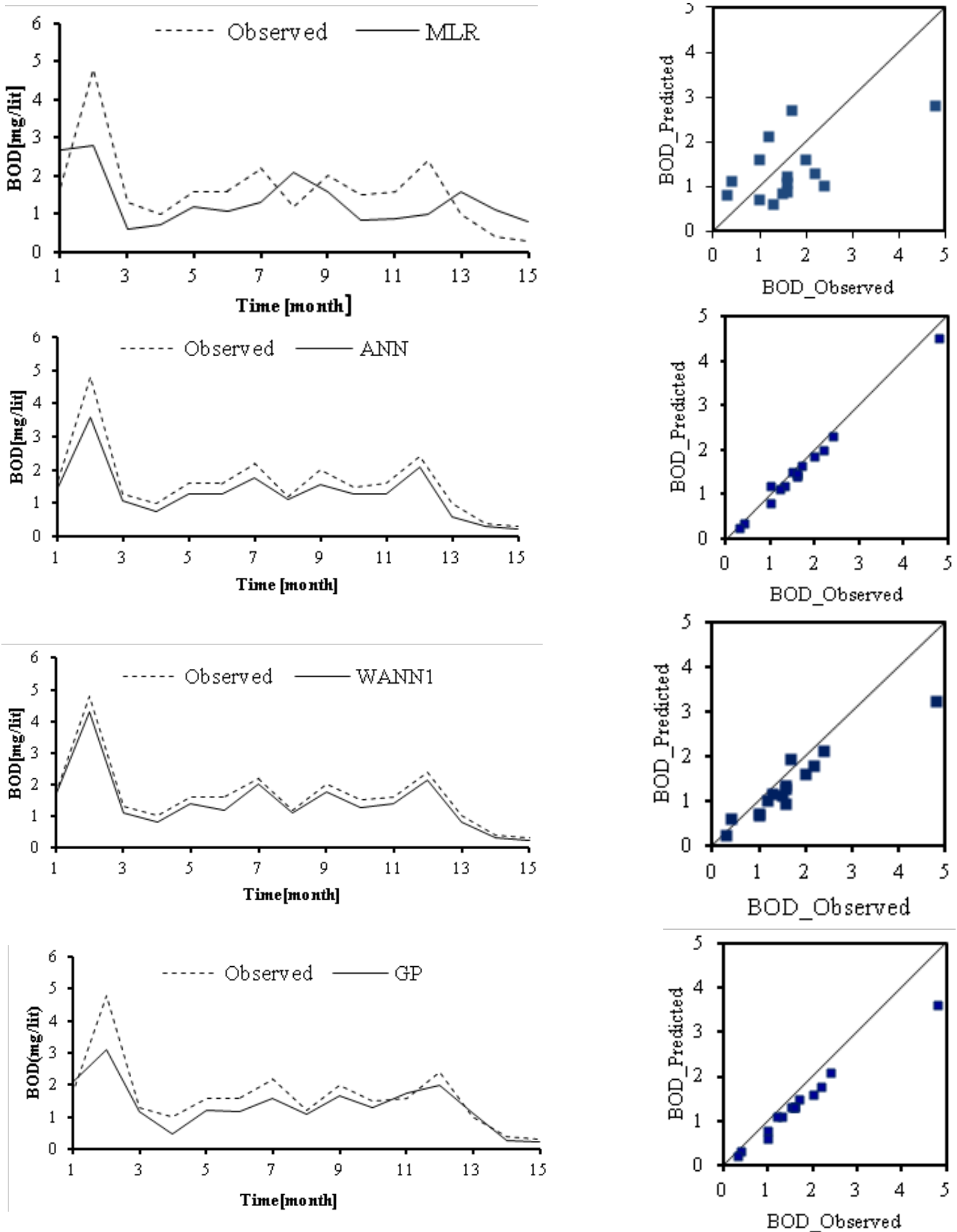
جدول ۴: نتایج مدل پیشنهادی اول ترکیب موجک

Table4. Results of the first proposed of wavelet combination

MAE (mg/l)	RMSE (mg/l)	E	ساختار شبکه عصبی	سطوح تجزیه	انواع موجک مادر
۱/۲۳	۰/۸۹	۰/۵۴	۲-۵-۱	۱	دابیچز ۲
۱/۱۸	۰/۸۴	۰/۶۱	۲-۷-۱	۲	
۰/۷۸	۰/۴۶	*۰/۸۱	۲-۳-۱	۳	
۱/۶۵	۱/۱	۰/۴۸	۲-۱۱-۱	۴	
۱/۹۸	۱/۴۵	۰/۳۵	۲-۹-۱	۵	
۱/۱۸	۰/۷۸	۰/۶۶	۲-۶-۱	۱	دابیچز ۵
۱/۹۸	۱/۴۴	۰/۳۴	۲-۵-۱	۲	
۱/۱۵	۰/۷۵	۰/۶۸	۲-۹-۱	۳	
۲/۱۱	۱/۸۶	۰/۲۳	۲-۱۰-۱	۴	
۱/۹۴	۱/۸۳	۰/۲۵	۲-۷-۱	۵	
۰/۷۷	۰/۵۸	۰/۷۶	۲-۶-۱	۱	میر
۱/۱۸	۰/۷۸	۰/۶۵	۲-۴-۱	۲	
۱/۱۷	۰/۷۸	۰/۶۶	۲-۹-۱	۳	
۱/۸۱	۱/۲۳	۰/۴۳	۲-۸-۱	۴	
۱/۸۴	۱/۲۶	۰/۴۱	۲-۱۰-۱	۵	
۲/۳۷	۲/۰۸	۰/۱۲	۲-۵-۱	۱	هار
۱/۸۴	۱/۲۶	۰/۴۱	۲-۷-۱	۲	
۱/۹۸	۱/۴۵	۰/۳۳	۲-۴-۱	۳	
۱/۹۸	۱/۴۵	۰/۲۱	۲-۹-۱	۴	
۲/۳۴	۱/۹۵	۰/۱۴	۲-۱۱-۱	۵	

توانسته است ساز و کارهایی را برای جبران تغییرات به وجود آمده در طول زمان ارائه دهد. به خصوص مقادیر بیشینه و کمینه را به خوبی شناسایی کرده ولی در پیش‌بینی مقدار آن‌ها اندکی خطا داشته است. در این مدل نقاط اطراف خط نیم‌ساز نوسان کرده‌اند که نشان می‌دهد این مدل از پارامترهای بیشتری برای تنظیم وزن‌ها برخوردار بوده است. مدل شبکه عصبی با بهینه کردن وزن‌ها تا حد قابل قبولی توانسته این پارامتر را پیش‌بینی کند. این حالت در نمودار مربوط به موجک به دقت بیشتری می‌رسد به‌طوری‌که نقاط بر خط نیم‌ساز انطباق خوبی دارند. مدل عصبی-موجکی نه تنها قادر به تشخیص تغییرات سری زمانی بوده بلکه نقاط شکستگی، بیشینه و کمینه را نیز به خوبی پیش‌بینی کرده است. با بررسی نقاط شکستگی نمودارهای شبکه عصبی و مدل ترکیبی می‌توان دریافت که کارایی موجک باعث شده که در قله‌ها با تشخیص نویز و استفاده از الگوی مناسب، ورود داده‌های بزرگ شناسایی شده و پیش‌بینی به سمت دقیق‌تر شدن پیش رود.

در مورد عوامل دوره‌ای نظیر دبی، دما و سرعت آب، به دلیل اینکه محدوده تغییرات مشخصی دارند، منطقی است که باید تأثیرگذاری آن‌ها را در روند کلی سیگنال اصلی جستجو نمود. این رفتار فرکانس پایین، قابل بازسازی توسط زیر سری تقریبات (۲۳) می‌باشد. شکل ۶، نمودارهای مدل سازی BOD توسط سه مدل در دوره آزمون را در مقایسه با مقادیر مشاهده شده نشان می‌دهد. عملکرد ضعیف روش رگرسیون خطی در شکل ۶ با پراکندگی داده‌ها در اطراف خط نیم‌ساز دیده می‌شود. هرچه این نقاط به نیم‌ساز نزدیکتر باشند، پیش‌بینی بهتر انجام شده است. به نظر می‌رسد روش رگرسیون مدل مقادیر میانگینی را ارائه داده و در هیچ یک از نوسان‌ها و پیک‌ها با مقادیر مشاهداتی تطابق ندارد. در واقع مدل رگرسیون کلیات روند تغییرات سری زمانی را با چند گام تأخیر دنبال کرده و نتوانسته است جزئیات را شبیه‌سازی کند که این موضوع به ماهیت خطی این مدل بر می‌گردد. این روش قادر به تشخیص روابط پیچیده بین داده‌های سری زمانی نبوده و پیش‌بینی با خطای نسبتاً بالایی همراه است. در صورتیکه شبکه عصبی بهتر



شکل ۶: مقایسه مدل‌های پیش‌بینی رگرسیون، شبکه عصبی، برنامه‌ریزی ژنتیک و مدل پیشنهادی اول

Fig.6. Compare of prediction models include of regression, neural network, genetic programming and first proposed model



۳-۲- مدل پیشنهادی دوم

بهترین نتیجه در روش ترکیبی اول با موجک دایچیز ۲ به دست آمد، لذا در مدل دوم این موجک مینا قرار گرفته و در ده سطح تجزیه شده است. هر یک از این ده سطح، ارتباطهای تناوبی و همبستگی بین BOD و تأخیرهای زمانی آن را به صورت ماهانه رمزگشایی می کنند. هدف اصلی تجزیه جدا نمودن اختلالات از نشانه پایه به وسیله تبدیل موجکی به منظور دست یافتن به هندسه و روند گروه و در نتیجه امکان پیش بینی بهتر آن می باشد. گروه های زمانی پس از حذف اختلالات به زیرسری هایی ایستا تبدیل می شوند که به علت افزایش حافظه سری زمانی دارای همبستگی بالاتری می باشند. با توجه به جدول ۵، بالاترین ضریب

همبستگی بین زیرسری های BOD با یک و سه تأخیر مربوط به  $DDW_1$ ،  $DDW_2$ ،  $DDW_3$ ،  $DDW_6$  می باشد. بنابراین جمع این زیرسری ها، ترکیب  $BDDW_t$  و  $BDDW_{t-6}$  را ایجاد می کند. سایر ترکیب های زیر نیز به این ترتیب ایجاد می شوند. این ترکیب ها برای موجک مادر دایچیز ۲ و در ۵ سطح با تعداد نرون های لایه پنهان ۱ تا ۱۵ (تعداد ۷۵ آنالیز) انجام می گیرد.

- ۱)  $BOD_t$
- ۲)  $BOD_t, BOD_{t-3}$
- ۳)  $BOD_t, BOD_{t-6}$
- ۴)  $BOD_t, BOD_{t-3}, BOD_{t-6}$
- ۵)  $BOD_t, BOD_{t-3}, BOD_{t-4}, BOD_{t-6}$

جدول ۵: ضرایب همبستگی زیرسری های تجزیه شده با BOD

Table 5. Correlation coefficients of BOD decomposed subseries with BOD

ضریب همبستگی بین BODt و $DDW(i)_t$ ۱-	ضریب همبستگی بین BODt و $DDW(i)_t$ ۳-	ضریب همبستگی بین BODt و $DDW(i)_t$ ۴-	ضریب همبستگی بین BODt و $DDW(i)_t$ ۶-	زیرسری های گسسته شده BOD
۰/۳۱	۰/۲۳	۰/۱۸	۰/۲۷	۱ DDW
۰/۱۸	۰/۰۹	۰/۰۱	۰/۲۱	۲ DDW
۰/۲۴	۰/۱۴	۰/۱۱	۰/۱۹	۳ DDW
۰/۱۳	۰/۰۵	۰/۱۵	۰/۱۶	۴ DDW
۰/۱۴	۰/۱۳	۰/۰۴	۰/۰۸	۵ DDW
۰/۱۷	۰/۲۷	۰/۲۳	۰/۱۴	۶ DDW
-۰/۲۱	-۰/۱۷	۰/۰۵	-۰/۱۲	۷ DDW
-۰/۱۱	۰/۰۴	۰/۰۹	-۰/۰۹	۸ DDW
۰/۰۶	-۰/۱۸	۰/۱۲	-۰/۰۷	۹ DDW
-۰/۱۸	۰/۰۵	-۰/۰۱	۰/۱۶	۱۰ DDW
۰/۰۳۴	۰/۰۲۴	۰/۰۱۷	۰/۰۴۳	.DDW App

جدول ۶: ضرایب همبستگی بین BDDW، با سیگنال اصلی BOD

Table 6. Correlation coefficients between BDDW and original signal of BOD

ضریب همبستگی سری های زمانی با سری اصلی BOD			سری های زمانی
همه داده ها	مجموعه آزمون	مجموعه آموزش	
۰/۷۲	۰/۶۸	۰/۷۵	BDDWt
۰/۵۵	۰/۵۲	۰/۵۵	BDDWt-۳
۰/۵۵	۰/۵۵	۰/۴۸	BDDWt-۴
۰/۷	۰/۷۱	۰/۶۸	BDDWt-۶
۰/۷	۰/۶۸	۰/۷۱	BODt
۰/۵۴	۰/۵۱	۰/۵۴	BODt-۳
۰/۵۱	۰/۵۶	۰/۴۸	BODt-۴
۰/۶۸	۰/۶۶	۰/۶۵	BODt-۶

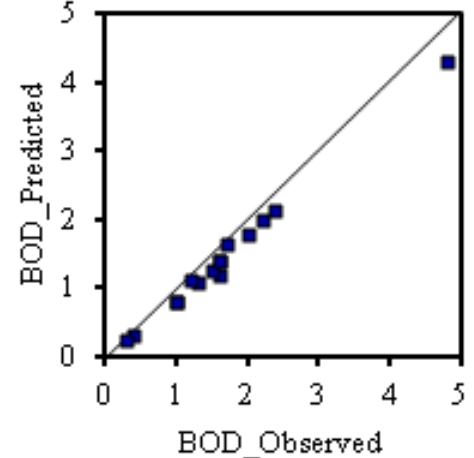
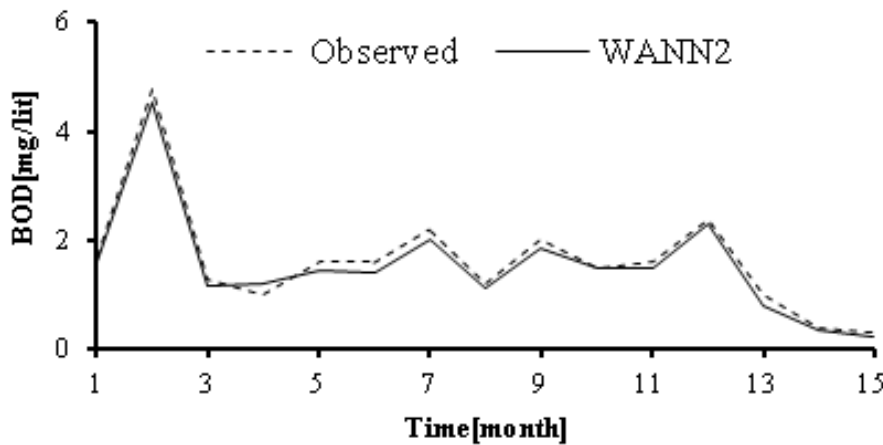
با توجه به جدول ۶ ضرایب همبستگی بین BOD و سیگنال بدون پیش پردازش BOD با تأخیرهای یک، سه، چهار و شش ماهه به ترتیب  $0/68, 0/51, 0/56, 0/66$  می باشد. در حالیکه نویززدایی موجک این ضریب را به  $0/68, 0/52, 0/55, 0/71$  بهبود داده است. به همین علت یک مدل ترکیبی که از سیگنال های چند مقیاسه مؤثرتر استفاده می کند، می تواند پیش بینی دقیق تری را ارائه دهد.

مطابق جدول ۷ بهترین نتیجه در مدل دوم با ضریب تبیین  $0/81$  به دست آمد.

جدول ۷: نتایج مدل پیشنهادی دوم

Table 7. Results of the second proposed model

WANN (BOD)					انواع مدلها
۵	۴	۳	۲	۱	ترکیبها
۹	۵	۷	۵	۳	تعداد نرونهای لایه پنهان
۰/۶۲	۰/۷۱	۰/۸۳	۰/۷۶	۰/۶۸	E
۰/۸۴	۰/۶۹	۰/۴۴	۰/۵۸	۰/۷۴	MAE (mg/l)
۱/۱۵	۱/۱۱	۰/۷۲	۰/۷۷	۰/۸۷	RMSE (mg/l)



شکل ۷: نمودار مقادیر پیش‌بینی و مشاهده شده برای مدل ترکیبی پیشنهادی دوم

Fig.7. Graph of predicted and observed values for the second proposed hybrid model

در افزایش ضریب کارایی و کاهش میزان خطای داده‌های محاسباتی به مشاهداتی، موفقیت‌آمیز بوده است. مدل ترکیبی دوم، دقت بالاتری را نسبت به سایر روش‌ها نشان می‌دهد.

۳-۳- توانایی تخمین پیک‌ها

جدول ۸ توانایی پیش‌بینی مقادیر بیشینه در دوره آزمون را توسط سه مدل نشان می‌دهد. استفاده هم‌زمان شبکه‌های عصبی و تبدیل موجک

جدول ۸: پیش‌بینی مقادیر بیشینه بازه آزمون توسط چهار مدل برای BOD

Table 8. predicted of maximum values in test period by four models for BOD

WANN <sub>۲</sub>	WANN <sub>۱</sub>	GP	ANN	MLR	اندازه‌گیری شده	تاریخ	ردیف
۲/۱۳	۲/۰۵	۲/۲	۱/۹	۲/۷	۱/۷	۸۹ خرداد	۱
۴/۰۲	۳/۹	۲/۹	۳/۲	۲/۸	۴/۸	۸۹ تیر	۲
۲	۱/۹	۱/۶	۱/۷۸	۱/۳	۲/۲	۸۹ آذر	۳
۱/۸	۱/۷۸	۱/۶۸	۱/۵۸	۱/۶	۲	۸۹ بهمن	۴
۲/۳	۲/۰۳	۲	۲/۱	۱	۲/۴	۹۰ اردیبهشت	۵
۱۲/۳	۱۱/۸۶	۱۰/۴۸	۱۰/۵۶	۱۰/۳	۱۳/۱		جمع مقادیر
%۶	%۹	%۲۰	%۱۹	%۲۱			خطا (%)
۰/۱۷	۰/۲۵	۰/۷۳	۰/۶۲	۰/۸۸			RMSE

مقدار غلظت BOD آب رودخانه کرج در محل خروجی سد تنظیمی کرج بررسی شد. با مشاهده سری زمانی BOD، به نظر می‌رسید که این سری ماهیت دینامیکی دارد، لذا لحاظ کردن تغییرات زمانی به تنهایی دقت قابل

۴- نتیجه گیری

در این تحقیق کاربرد مدل‌های شبکه عصبی، برنامه‌ریزی ژنتیک، ترکیب شبکه عصبی-موجک و رگرسیون خطی چندمتغیره در پیش‌بینی

- Neural-Wavelet Model for Prediction of Ligvanchai Watershed Precipitation, *Engineering Application of Artificial Intelligence*, Vol. 22, No. 3, pp.146-477, 2009.
- [6] T., Rajaei; Wavelet-ANN combination model for prediction of daily suspended sediment load in rivers, *Science of The Total Environment*, Vol. 409, No. 15, pp. 2917- 2928, 2011.
- [7] T., Rajaei; V., Nourani; M., Zounemat-Kermani; O., Kisi; River Suspended Sediment Load Prediction: Application of ANN and Wavelet Conjunction Model, *Journal of Hydrologic Engineering*, Vol. 16, pp.613-627, 2011.
- [8] R.M., Singh; Wavelet-ANN Model for Flood Events, *Advances in Intelligent and Soft Computing*, Vol. 131, pp.165-175, 2012.
- [9] Xu ., Longqin; Liu., Shuangyin; Study of short-term water quality prediction model based on wavelet neural network, *Mathematical and Computer Modelling*, Vol. 58, pp.807-813, 2013.
- [10] V., Nourani; A., Hosseini Baghanam; A., Adamowski; M., Gebremichael; Using self-organizing maps and wavelet transforms for space–time pre-processing of satellite precipitation and run off data in neural network based rainfall–runoff modeling, *Journal of Hydrology*, Vol. 476, pp.228-243, 2013.
- [11] L., Shuangyin; A hybrid WA-CPSO-LSSVR model for dissolved oxygen content prediction in crab culture, *Engineering Applications of Artificial Intelligence*, Vol, 29, pp.114-124, 2014.
- [12] M., Alizadeh; M., Kavianpour; Development of wavelet-ANN models to predict water quality parameters in Hilo Bay, *Marine Pollution Bulletin*, Vol. 98, No. 1-2, pp. 171-178, 2015.
- [13] Z., Jamshidzadeh; M.R., Alavi moghadam; Assessment of surface water quality based on CWQI index, *Proceedings of the National Conference on Flow and Water Pollution*, pp.1-10, 1385 (In Persian).
- [14] S., Haykin; Neural Networks: a comprehensive foundation, *MacMillan, New York*, 1994.
- [15] J.G., Han; W.X., Ren; Z.S., Sun; Wavelet packet based damage identification of beam structures, *International Journal of Solids and Structures*, Vol. 42, No. 26, pp. 6610-6627, 2005.
- [16] J., Adamowski; F. H., Chan; A wavelet neural network conjunction model for groundwater level forecasting, *Journal of Hydrology*, Vol. 407, No. 1-4, pp.28- 40, 2011.
- قبولی در این تحقیق نداشته و بایستی دینامیک محیط نیز در مدل‌ها توسط آنالیز موجکی وارد شود. نتایج تحقیق نشان داد که موجک با درک مکانیسم سری زمانی قادر است تفسیری را از ساختار سری زمانی مذکور و اطلاعات مهم و کلیدی را از حافظه آن استخراج کند. مطابق ارزیابی‌های انجام شده، روش رگرسیون دقت لازم را برای پیش‌بینی نداشت و به خصوص در پیش‌بینی نقاط بیشینه و کمینه، بسیار ضعیف عمل کرد، در حالیکه روش‌های دیگر پیش‌بینی مناسبی داشتند. مدل ترکیبی عصبی-موجک در قالب دو مدل پیشنهادی ارائه شد. مدل اول بر اساس تأخیر سیگنال اصلی و جمع جبری همه زیرسری‌های حاصل از آن و مدل دوم بر اساس جمع جبری زیر سری‌های مؤثر و تأخیر هر زیر سری انجام گرفت. هر دو مدل ترکیبی نتایج بهتری را نسبت به مدل‌های عصبی، ژنتیک و رگرسیون داشتند. مدل پیشنهادی دوم توانست، کمترین خطا و بالاترین دقت را ارائه دهد. این مدل توانست دقت ضریب کارائی نش را نسبت به روش‌های رگرسیون، شبکه عصبی، برنامه‌ریزی ژنتیک و مدل ترکیبی اول به ترتیب از ۰/۷۳، ۰/۵۳ و ۰/۸۱ به ۰/۸۳ افزایش دهد و میانگین قدرمطلق خطا را از ۱/۸۷، ۰/۶۵ و ۰/۹۱ به ۰/۴۶ و ۰/۴۴ کاهش دهد. پیشنهاد می‌شود در تحقیقات بعدی به منظور کاهش اغتشاشات سیگنال از سری‌های زمانی روزانه یا هفتگی در بازه زمانی طولانی‌تری استفاده شود. همچنین از پارامتر دما و DO به عنوان سیگنالی تأثیرگذار در پیش‌بینی BOD در ترکیبات ورودی به مدل شبکه عصبی استفاده شود.

## مراجع

- [1] A., Najah; A., Elshafie; O., Karimi; O., Jaffer; Prediction of Johor river water quality parameters using artificial neural networks, *European Journal of Scientific Research*, Vol. 28, pp. 422-435, 2009.
- [2] T., Rajaei; A., Mirbagheri; The suspended load model of rivers using artificial neural networks, *Journal of Ferdowsi university of Mashhad engineering faculty*, Vol. 21, pp. 1-11, 2009 () .
- [3] S.E., Kim; I.W., Seo; Artificial Neural Network ensemble modeling with conjunctive data clustering for water quality prediction in rivers, *Journal of Hydro-environment Research*, Vol. 9, No. 3, pp.1-15, 2015.
- [4] O., Makarynsky; D., Makarynska; M., Rayson; S., Langtry; Combining deterministic modelling with artificial neural networks for suspended sediment estimates, *Applied Soft Computing*, Vol. 35, pp. 247–256, 2015.
- [5] V., Nourani; M.T., Alami; M.H., Aminfar; A Combined

برای ارجاع به این مقاله از عبارت زیر استفاده کنید:



Please cite this article using:

T. Rajaei, H. Jafari, R. Rahimi, "Performance Improvement of Biological BOD in Rivers based on De-noising Comparison Wavelet-ANN Conjunction, GP, ANN and MLR Methods (Case Study: Karaj Dam Outlet Station)". *Amirkabir J. Civil Eng.*, 49(2) (2017) 273-283.

DOI:10.22060/ceej.2016.710

