

مقایسه روش‌های برنامه‌ریزی بیان ژن، سری زمانی غیرخطی، خطی و شبکه عصبی مصنوعی در تخمین دبی روزانه (مطالعه موردی: رودخانه کارون)

رضا زمانی^{۱*} - فرشاد احمدی^۲ - فریدون رادمنش^۳

تاریخ دریافت: ۱۳۹۲/۱۲/۲

تاریخ پذیرش: ۱۳۹۳/۶/۱۰

چکیده

امروزه پیش‌بینی جریان روزانه رودخانه‌ها از مباحث مهم در هیدرولوژی و منابع آب می‌باشد و می‌توان از نتایج الگوبندی جریان روزانه رودخانه در مدیریت منابع آب، خشکسالی‌ها و سیلاب‌ها استفاده کرد. با توجه به اهمیت این موضوع، در این پژوهش با استفاده از روش‌های سری زمانی خطی، غیرخطی و الگوهای هوش مصنوعی (شبکه عصبی و برنامه‌ریزی بیان ژن) به الگوبندی جریان روزانه رودخانه کارون در محل ایستگاه آب‌سنجی ارمند طی دوره آماری (۱۳۹۰-۱۳۶۰) پرداخته شده است. حوضه بالادست ایستگاه ارمند از جمله زیرحوضه‌های اصلی حوضه کارون شمالی است که شامل چهار زیرحوضه ونک، کارون میانی، بهشت آباد و کوه‌رنگ است. نتایج این پژوهش نشان داده است که الگوهای هوش مصنوعی دارای برتری نسبت به الگوی غیرخطی و خطی سری زمانی در الگوبندی جریان روزانه رودخانه کارون می‌باشند. همچنین الگوبندی و مقایسه الگوهای هوش مصنوعی نیز نشان داد که روش برنامه‌ریزی بیان ژن دارای معیارهای ارزیابی مناسب‌تری نسبت به روش شبکه عصبی مصنوعی می‌باشد.

واژه‌های کلیدی: الگوبندی، هوش مصنوعی، دبی روزانه، رودخانه کارون

مقدمه

نسبت به جنبه‌های دیگر به فرآیند خطی نزدیک‌تر باشند (۲۹). الگوی دوخطی در این راستا توسط گرانگر و آندرسون (۱۷) معرفی و پژوهش‌های فراوانی پس از آن بر روی این الگو انجام شده است. سوبارائو و گابری (۲۹) بر روی برخی خصوصیات و کاربردهای این الگو تحقیق و نمایش ماتریسی و فضایی آن را ارائه نمودند. دای و بیلارد (۷) مشکل تخمین پارامترهای الگوی دوخطی را مورد توجه قرار داده و روشی به‌نام برآورد حداکثر درست‌نمایی شرطی با فرض ایستا بودن سری و با استفاده از روش نیوتن رافسون^۴ ارائه نمودند. لیفشیتس (۲۳) منشا تغییرات الگوی دوخطی را با ماهیت غیرخطی ضعیف بررسی و به روابط و معادلات ریاضی حاکم بر این شرایط دست یافت. چند پژوهش مهم سری زمانی غیرخطی در ادامه معرفی می‌شود.

وانگ و همکاران (۳۰) برای برازش واریانس و میانگین روزانه جریان رودخانه زرد در چین از ترکیب الگوهای ARMA^۵ و

الگوبندی بسیاری از فرآیندهای هیدرولوژیکی کاری سخت ولی با اهمیت است. زیرا متغیرهای مختلفی دستگاه پیچیده‌ای را در این راستا با تاثیرهای متقابل ایجاد می‌کنند. لذا سامانه‌هایی برای پیش‌بینی فرآیندهای هیدرولوژیکی به صورت مکانی یا زمانی در مدیریت منابع آب نیاز است. گوویندراجو (۱۶) الگوهای مرسوم هیدرولوژی را به الگوهای ریاضی-فیزیکی، ژئومورفولوژی و تجربی تقسیم کرده است. الگوهای سری زمانی، ریاضی فیزیکی محسوب می‌شوند و توانایی زیادی در الگوبندی پدیده‌های خطی و غیرخطی دارند. الگوهای سری زمانی از دو بخش اصلی، شامل مولفه غیرقطعی (تصادفی) و مولفه قطعی (جبری) تشکیل شده‌اند (۲۶). الگوهای خطی رایج‌ترین الگوی سری زمانی در هیدرولوژی و منابع آب است (۲۵). بسیاری از فرآیندهای مربوط به سامانه‌های طبیعی نسبت به زمان غیرخطی هستند. جنبه‌های خاصی از این سامانه‌ها می‌توانند

۱، ۲ و ۳- به ترتیب دانشجویان دکتری و دانشیار گروه منابع آب، دانشکده مهندسی

علوم آب، دانشگاه شهید چمران اهواز

*- نویسنده مسئول: (Email: reza.zamani65@yahoo.com)

4- Newton-Raphson

5- Autoregressive Moving Average

عصبی فازی^۴ را در روندیابی سیلاب رودخانه قزل ایرماق ترکیه ارزیابی کردند. نتایج پژوهش آنها نشان داد که روش برنامه‌ریزی ژنتیک هیدروگراف خروجی را با دقت بیشتری شبیه‌سازی می‌کند. قبادیان و همکاران عملکرد روش برنامه‌ریزی بیان ژن^۵ را در روندیابی روندیابی سیلاب رودخانه زنگمار در مقایسه با روش موج دینامیکی بررسی کردند. نتایج روندیابی شش آب‌نمود سیل (۸۸-۸۴) نشان داد که الگوی برنامه‌ریزی بیان ژن قادر است با دقت بیشتری حجم آب‌نمود خروجی را پیش‌بینی کند. اما الگوی موج دینامیک در خصوص دبی اوج و زمان وقوع آن برتری دارد.

امروزه علاقه‌ی فزاینده‌ای در توسعه نظری سامانه‌های پویای هوشمند که مبتنی بر داده‌های تجربی هستند ایجاد شده است. شبکه‌های عصبی مصنوعی جزء این دسته‌اند که دانش یا قانون نهفته در ورای این اطلاعات را با پردازش روی داده‌های تجربی به ساختار شبکه منتقل می‌کنند. روش‌های هوشمند شبکه‌های عصبی مصنوعی (ANNs) نیز به‌منظور پیش‌بینی جریان استفاده شده است (۲۲). دوگان و همکاران (۹) برای پیش‌بینی جریان روزانه در پژوهشی از دو روش شبکه عصبی مصنوعی و اتورگرسیون استفاده کرده‌اند. فیرات (۱۳) برای پیش‌بینی جریان روزانه رودخانه در حوضه‌ای در ترکیه از سه روش شبکه عصبی پیش‌خور، روش عصبی فازی و روش شبکه عصبی تعمیم یافته رگرسیونی استفاده کردند. از جمله پژوهش‌گران دیگری که از شبکه‌های عصبی در الگوبندی رواناب رودخانه استفاده کرده‌اند می‌توان به کیشی (۲۲)، باره و همکاران (۵)، لوهانی و همکاران (۲۴) اشاره کرد.

با توجه به اهمیت بحث الگوبندی جریان روزانه رودخانه، نقش آن در مدیریت و برنامه‌ریزی منابع آب، مقایسه کاربرد روش‌های مختلف و همچنین مقایسه عملکرد سری زمانی خطی و غیرخطی، در این پژوهش با استفاده از الگوهای برنامه‌ریزی بیان ژن، سری‌های زمانی خطی، غیرخطی و شبکه عصبی به الگوبندی دبی روزانه رودخانه کارون در محل ایستگاه آب‌سنجی ارمند واقع در حوضه کارون شمالی پرداخته شده است.

مواد و روش‌ها

منطقه مورد مطالعه

حوضه آبریز بالادست ایستگاه آب‌سنجی ارمند یکی از زیرحوضه‌های اصلی حوضه کارون شمالی است. حوضه آبریز کارون شمالی در محدوده جغرافیایی ۳۴' - ۴۹° تا ۴۷' - ۵۱° طول شرقی و ۱۸' - ۳۱° تا ۴۰' - ۳۲° عرض شمالی قرار دارد.

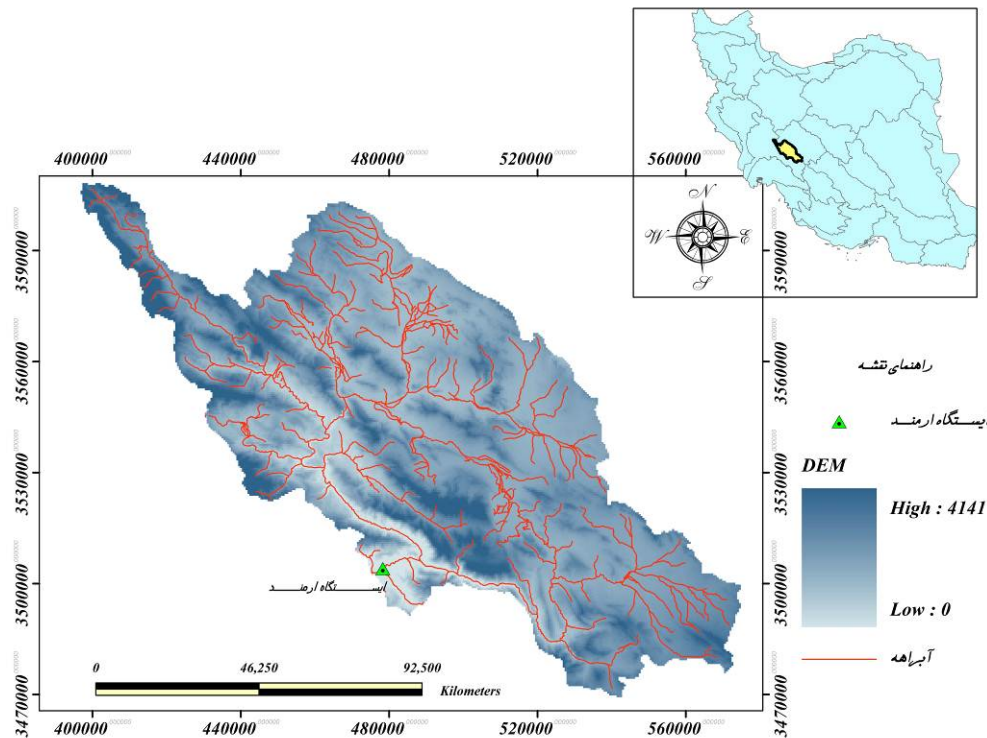
GARCH^۱ استفاده کردند. نتایج این پژوهش نشان داد که الگوی ARMA-GARCH نتایج قابل قبولی در الگوبندی سری روزانه جریان رودخانه مورد مطالعه ارائه کرده است. کایادو (۶) عملکرد الگوهای سری‌های زمانی یک متغیره را در پیش‌بینی میزان آب مصرفی (در مقیاس روزانه و هفتگی اسپانیا) در طی دوره آماری (۲۰۰۶-۲۰۰۱) بررسی کرد. الگوهای ARIMA و GARCH بر سری داده‌های مشاهداتی برازش و کارایی این دو ارزیابی و تایید شد. ناوه و همکاران (۲۵) الگوی دوخطی را برای پیش‌بینی جریان ماهانه رودخانه‌های شهرچای و باراندوزچای ارومیه استفاده کردند. نتایج نشان از دقت بالای این الگو در مقایسه با الگوهای خطی است. همچنین کارایی الگوی غیرخطی دوخطی در پیش‌بینی جریان رودخانه تایید شد. خلیلی و همکاران (۲۰) جریان روزانه رودخانه شهرچای ارومیه را با الگوی غیرخطی تلفیقی BL-ARCH پیش‌بینی کردند. نتایج نشان داد که الگوی تلفیقی عملکرد بسیار بهتری نسبت به الگوی دوخطی داشته و مقدار خطا را به میزان قابل توجهی کاهش می‌دهد.

برنامه‌ریزی ژنتیک^۲ یک روش دیگر پرکاربرد در منابع آب و هیدرولوژی می‌باشد. این الگو یک فن خودکار است که راه حل مسئله را با برنامه‌نویسی ارائه می‌کند. الگوریتم‌های تکاملی که برنامه‌ریزی ژنتیک نیز عضوی از آنها هستند توانایی الگوبندی فرآیندهای کاملاً غیرخطی و پویا را دارند. بنا به اهمیت موضوع، تاکنون پژوهش‌گران مختلفی در سراسر جهان اقدام به مطالعه جریان رودخانه‌ها نموده‌اند. خو و همکاران (۲۱) رواناب ساعتی حوضه‌ای در فرانسه را با روش برنامه‌ریزی ژنتیک پیش‌بینی کردند. بررسی نتایج این پژوهش نشان داد که برنامه‌ریزی ژنتیک توانایی پیش‌بینی دقیق رواناب را در تمام فواصل زمانی (به‌ویژه فواصل کوتاه‌تر از زمان تمرکز حوضه) را دارد. همچنین نتایج با مطالعه انجام شده دیگر که در آن روش‌های خودهمبستگی و صافی کالمن استفاده شده بود مقایسه و مشخص شد که برنامه‌ریزی ژنتیک ابزار بهتری برای پیش‌بینی رواناب است. همچنین روش برنامه‌ریزی ژنتیک مطابق پژوهش آیتک و همکاران (۴) یک روش مناسب و علمی در پیش‌بینی روابط بارش-رواناب است. گوون (۱۸) روش برنامه‌ریزی ژنتیک خطی را برای پیش‌بینی دبی روزانه رودخانه شویل کیل در ایالات متحده استفاده کرد. او نتایج حاصل را با دو الگو از روش‌های شبکه عصبی مقایسه کرد. وی نشان داد که هر دو روش نتایج قابل دارند. ولی دقت روش برنامه‌ریزی ژنتیک خطی بیشتر از روش شبکه عصبی است. قربانی و همکاران (۱۵) عملکرد سه روش برنامه‌ریزی ژنتیک، شبکه عصبی مصنوعی^۳ و

- 1- Generalized Autoregressive Conditional Heteroskedasticity (GARCH)
- 2- Genetic Programming
- 3- Artificial Neural Network (ANN)

4- Neuro-fuzzy

5- Gen Expression Programming



شکل ۱- موقعیت حوزه مورد مطالعه و ایستگاه آبسنجی ارمند

سری‌های زمانی هیدرولوژیکی است و حالت خاصی از الگوی غیر-خطی دوخطی محسوب می‌شود شکل کلی آن مطابق رابطه ۱ است (۲۸).

$$\text{ARMA}(p,q): Z_t = \sum_{i=1}^p (\varphi_i \cdot Z_{t-i}) - \sum_{j=1}^q (\theta_j \cdot \varepsilon_{t-j}) + \varepsilon_t \quad (1)$$

الگوهای خطی سری زمانی در واقع بسط مرتبه اول سری‌های تیلور^۱ هستند. ایده اصلی الگوی دوخطی نیز غیرخطی بودن بسط مرتبه دوم سری تیلور است. شکل کلی الگوی دوخطی مطابق رابطه ۲ است (۲۸).

$$Z_t = \sum_{i=1}^p (\varphi_i \cdot Z_{t-i}) - \sum_{j=1}^q (\theta_j \cdot \varepsilon_{t-j}) + \sum_{i=0}^r \sum_{j=1}^s (\beta_{ij} \cdot Z_{t-i-j} \cdot \varepsilon_{t-j}) + \varepsilon_t \quad (2)$$

در رابطه بالا، Z_t سری زمانی مورد نظر، θ_j ضریب مدل MA و p ، q ، r و s اعداد صحیح مثبت می‌باشند که رسته یا مرتبه الگوی دوخطی را نشان می‌دهند. الگوی فوق در برخی منابع مانند فان و یائو (۱۰) به صورت الگوی $BL(p, q, r, s)$ نیز نمایش داده شده است. الگوی دوخطی در واقع الگوی ARMA (خطی) بسط داده شده که

عبارت غیرخطی $\sum_{i=0}^r \sum_{j=1}^s (\beta_{ij} \cdot Z_{t-i-j} \cdot \varepsilon_{t-j})$ به سمت راست آن

این حوزه از سمت شمال و شمال شرق به حوزه آبریز سد زاینده رود، از شمال غرب به حوزه رودخانه دز، از جنوب به حوزه آبریز رودخانه خرسان و از جنوب و غرب به بخش‌هایی از حوزه آبریز کارون بزرگ محدود می‌گردد. حوزه بالادست ایستگاه آبسنجی ارمند شامل زیرحوضه‌های بهشت آباد، ونک، کارون میانی و کوه‌رنگ است. این حوزه عمدتاً دارای رژیم برفی و مساحتی حدود ۱۱۰۰۰ کیلومترمربع است. متوسط بارندگی سالانه این حوزه حدود ۷۵۰ میلی‌متر، حداکثر ارتفاع این حوزه از سطح آب‌های آزاد ۴۱۴۱ متر و کمترین ارتفاع آن ۱۰۴۴ متر است. ایستگاه آبسنجی ارمند در خروجی این حوزه (در نزدیکی سد کارون ۴) بر روی رودخانه کارون در ارتفاع ۱۰۸۲ متری و در عرض جغرافیایی $31^{\circ} - 40'$ شمالی و طول جغرافیایی $50^{\circ} - 46'$ شرقی با میانگین آبدهی سالانه حدود ۱۱۰ مترمکعب بر ثانیه قرار دارد. شکل ۱ موقعیت منطقه مورد مطالعه در کشور، حوزه کارون بزرگ و موقعیت ایستگاه آبسنجی ارمند را نشان می‌دهد.

مبانی نظری

الگوی دوخطی (Bilinear)

الگوی ARMA رایج‌ترین الگوی کاربردی در پیش‌بینی

1- Taylor Series

$$\varepsilon_t = Z_t - \left(\sum_{i=1}^p \varphi_i Z_{t-i} - \sum_{j=1}^q \theta_j \varepsilon_{t-j} + \sum_{i=1}^m \sum_{j=1}^s \beta_{ij} Z_{t-i} \varepsilon_{t-j} \right) \quad (۶)$$

قابل ذکر است که در الگوبندی با روش سری زمانی، در پژوهش حاضر برای تخمین پارامترها از ۸۰ درصد داده (دوره آماری ۱۳۸۵-۱۳۶۰) استفاده و ۲۰ درصد (دوره آماری ۱۳۹۰-۱۳۸۵) نیز برای آزمون الگو انتخاب شد.

برنامه‌ریزی بیان ژنتیک

روش برنامه‌ریزی بیان ژن جزو روش‌های الگوریتم گردشی محسوب می‌شود که مبنای آنها براساس نظریه تکامل داروین استوار است. الگوریتم‌های یاد شده اقدام به تعریف یک تابع هدف در قالب معیارهای کیفی نموده و سپس تابع یاد شده را برای مقایسه جواب‌های مختلف حل مسئله در یک فرآیند گام به گام تصحیح ساختار داده‌ها به کار می‌گیرند و در نهایت، جواب مناسب را ارائه می‌نمایند. تفاوت اساسی موجود بین الگوریتم ژنتیک و برنامه‌ریزی بیان ژن به طبیعت هریک از افراد برمی‌گردد، به نحوی که افراد در الگوریتم ژنتیک ردیف‌های خطی با طول ثابت می‌باشند (کروموزم‌ها) ولی در برنامه‌ریزی بیان ژن، همان شاخه‌های مجزا می‌باشند. همچنین در برنامه‌ریزی بیان ژن بر ساختار درختی مجموعه‌ها تاکید می‌شود ولی الگوریتم ژنتیک، بر اساس سیستم ارقام دودویی عمل می‌نماید. فرآیند گام به گام حل یک مسئله با استفاده از برنامه‌ریزی بیان ژن شامل ۵ مرحله به شرح زیر می‌باشد (۱۵):

- ۱- انتخاب مجموعه ترمینال: که همان متغیرهای مستقل مسئله و متغیرهای حالت سامانه می‌باشند. ۲- انتخاب مجموعه توابع: که شامل عملگرهای حسابی، توابع آزمون و توابع بولی می‌باشد. ۳- شاخص اندازه‌گیری دقت مدل که بر مبنای آن می‌توان مشخص نمود که توانایی مدل در حل یک مسئله خاص تا چه اندازه می‌باشد. ۴- مولفه‌های کنترل: مقادیر مولفه‌های عددی و متغیرهای کیفی که برای کنترل اجرای برنامه‌ها استفاده می‌شوند. ۵- شرط توقف اجرای برنامه: که معیاری برای حصول نتیجه و توقف اجرای برنامه می‌باشد. در مطالعه حاضر از برنامه GeneXpro Tools برای توسعه و اجرای مدل‌های مبتنی بر برنامه‌ریزی ژنتیک استفاده شده است. برنامه مذکور بر اساس برنامه‌ریزی صریح ژنتیک استوار است (۱۲). برنامه‌ریزی بیان ژن ویرایش جدیدی از برنامه‌ریزی ژنتیک می‌باشد که به استنتاج برنامه‌های رایانه‌ای با اندازه‌ها و شکل‌های مختلف می‌پردازد. یکی از نقاط قوت برنامه‌ریزی ژنتیک آن است که معیار تنوع ژنتیکی بسیار ساده بوده و لذا عملگرهای ژنتیک در سطح کروموزوم عمل می‌نمایند. همچنین یکی دیگر از نقاط قوت این روش، عبارت است از طبیعت منحصر به فرد چند ژنی بودن آن که زمینه ارزیابی مدل‌های پیچیده‌ای را که شامل چندین زیرمدل می‌باشند فراهم می‌آورد (۱۲). در فرآیند الگوبندی جریان رودخانه

اضافه شده است (۱). حاصل ضرب دو متغیر ε_{t-j} و Z_{t-i-j} در این عبارت سبب غیرخطی شدن الگوست. برازش یک الگوی سری زمانی غیرخطی دوخطی بر داده‌های اندازه‌گیری شده شامل دو مرحله مهم است: یکی تعیین مرتبه الگو شامل (p,q,r,s) و دیگری تخمین ضرایب φ, θ, β . تعیین مرتبه الگو با معیار آکائیکه اصلاح شده (AICC) مطابق رابطه ۳ است (۲ و ۳).

$$AICC(p,q) = nLn(\hat{\sigma}_e^2) + \frac{2(p+q+r+s)n}{(n-p-q-r-s)} \quad (۳)$$

که در رابطه بالا، n تعداد داده‌ها، p مرتبه الگو در بخش خود همبسته، q مرتبه الگو در بخش میانگین متحرک و $\hat{\sigma}_e^2$ واریانس خطاهاست. می‌توان ضرایب الگوی دوخطی را با داشتن (p, q, r, s) از روش استاندارد تابع درست نمایی تخمین زد. تابع لگاریتم درست‌نمایی (احتمالی) به صورت رابطه ۴ است. همچنین برای ساده‌سازی عملیات مشتق‌گیری و محاسبه ضرایب از این معادله لگاریتم گرفته شده است (۱۰):

$$l(\theta, \sigma^2) = -\frac{N-p'}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{t=p'+1}^T \hat{\varepsilon}_t(\theta)^2 \quad (۴)$$

که در رابطه بالا $p' = \max\{p, r\}$ ، N تعداد داده‌ها و مقادیر $\hat{\varepsilon}_{p'}(\theta), \hat{\varepsilon}_{p'+1}(\theta), \dots$ شرایط $q' = \max\{q, s\}$ و با استفاده از الگوی عمومی دوخطی مطابق با رابطه ۲ قابل محاسبه خواهند بود.

در این پژوهش برای بررسی نکویی برازش الگوهای خطی و غیرخطی منتخب، از آزمون پورت مانتو استفاده شد. بدین منظور ابتدا سری باقیمانده الگو به دست آمده و برای تاخیر مورد نظر، ضرایب خودهمبستگی سری محاسبه شد. سپس آماره پورت مانتو (Q) با رابطه ۵ محاسبه شد (۲۶):

$$Q = n \sum_{k=1}^L r_k^2 \quad (۵)$$

که در رابطه بالا، n تعداد داده، L تعداد تاخیر، r_k ضریب خود همبستگی مرتبه kام و ε_t سری باقیمانده الگو می‌باشد. آماره آزمون پورت مانتو دارای توزیع کی‌دو با درجه آزادی $(L-p-q)$ است و در صورتی که $Q \leq \chi_{(1-\alpha), (L-p-q)}^2$ باشد، مقادیر ε_t ناهمبسته بوده و الگوی نظیر آن کفایت لازم را دارد. سطح معنی‌داری نیز $\alpha = 0.05$ در نظر گرفته شد. همچنین در پژوهش حاضر برای الگوی دوخطی درجه آزادی به صورت $(L-p-q-r-s)$ و سری باقیمانده آن (ε_t) با معلوم بودن مرتبه و ضرایب الگو و با در نظر گرفتن مقدار اولیه صفر برای $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots, \varepsilon_{t-q}$ به شرح زیر محاسبه شد:

عصب‌ها یا نرون‌های مصنوعی اجزای اصلی یک شبکه عصبی هستند. الگوی ورودی به یک گره شبیه یک سلول زیستی است که می‌توان آن را با برداری با N مولفه به صورت $X=(x_1, x_2, \dots, x_n)$ نشان داد. مجموع حاصل ضرب ورودی‌ها در وزن‌های نظیرشان را می‌توان با کمیت اسکالر s نمایش داد:

$$s = \sum_{n=1}^N w_n \cdot x_n = W^T \cdot X \quad (8)$$

که در رابطه بالا، $W=(w_1, w_2, \dots, w_n)$ بردار وزن نرون‌ها است. کمیت s سپس وارد یک تابع غیرخطی f می‌شود تا خروجی را نتیجه دهد.

$$y = f(s) \quad (9)$$

تابع غیرخطی انتقال معمولاً به فرم سیگموئید به شکل زیر تعریف می‌شود:

$$f(s) = (1 + \exp(-s))^{-1} \quad (10)$$

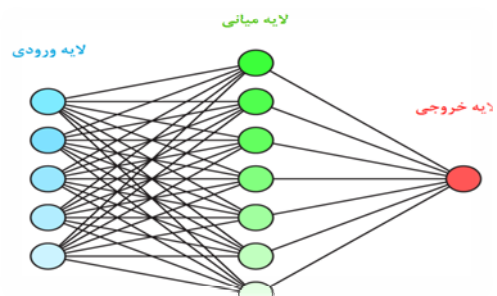
که خروجی y می‌تواند نتیجه الگو یا ورودی لایه بعدی در شبکه‌های چندلایه باشد. الگوریتم‌های مختلفی جهت محاسبه وزن‌های بهینه ارائه شده‌اند که در این میان الگوریتم "پس انتشارخطا" پرکاربردترین آن‌ها است. شبکه‌هایی که در این پژوهش به منظور الگوبندی دبی روزانه حوضه کارون در محل ایستگاه ارمند استفاده شدند از نوع شبکه‌های پرسپترون چند لایه است. این شبکه‌ها به صورت پیشرو عمل نموده و در ساختار آنها تا حد امکان طراحی بهینه با یک لایه میانی صورت می‌پذیرد. آموزش شبکه‌های پرسپترون چند لایه با الگوریتم آموزش پس انتشار خطا انجام شد. در شبکه‌های عصبی این پژوهش در تقسیم‌بندی الگوها به دو دسته آموزش و آزمون، به ترتیب ۸۰ و ۲۰ درصد در نظر گرفته شده است. در طی مرحله الگوسازی، ضرایب مربوط به خطای موجود در گره‌ها به صورت سعی و خطا تصحیح می‌شود که در اغلب موارد از نمایه خطای متوسط داده‌ها بهره برده می‌شود. این امر از طریق مقایسه خروجی‌های الگو با داده‌های ورودی مشاهداتی انجام می‌پذیرد.

قابل ذکر است که در پژوهش حاضر برای الگوبندی جریان روزانه ایستگاه ارمند با استفاده از برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی، داده‌های ۲۶ سال (از یک مهر ۱۳۶۰ تا سی و یک شهریور ۱۳۸۵) برای آموزش و ۵ سال (از یک مهر ۱۳۸۵ تا سی و یک شهریور ۱۳۹۰) به عنوان داده‌های تست انتخاب شدند. به طور کلی ۸۰ درصد داده‌ها برای آموزش و ۲۰ درصد نیز برای تست در نظر گرفته شد تا با الگوسازی سری‌زمانی جریان رودخانه مطابقت داشته باشد. در الگوبندی با استفاده از شبکه‌های عصبی مصنوعی از قانون آموزش لونیبرگ مارکوات، یک لایه پنهان و تعداد تکرار ۱۰۰۰ استفاده شده است. قابل ذکر است که، انتخاب جمعیت‌های اولیه تصادفی مختلف و تاثیرگذار در پدیده (که در برنامه‌ریزی بیان ژن به عنوان داده‌های

کارون در این پژوهش به صورت زیر عمل شده است: گام اول، انتخاب تابع برازش مناسب می‌باشد که در این مطالعه تابع جذر میانگین مربعات خطا به عنوان تابع برازش انتخاب گردید. گام دوم، انتخاب مجموعه متغیرهای ورودی و مجموعه توابع به منظور تولید کروموزوم‌ها می‌باشد. در مسئله حاضر مجموعه ترمینال‌ها متشکل از مقادیر جریان رودخانه با تاخیرهای زمانی می‌باشد. در این مطالعه از چهار عملگر اصلی که شامل $\{+, -, \times, \div\}$ و نیز توابع ریاضی $\{X^2, X^3, \sqrt[3]{X}, \sqrt{X}, \ln(X)\}$ استفاده به عمل آمد. گام سوم، شامل انتخاب ساختار و معماری کروموزوم‌ها می‌باشد. اندازه طول هر راس ۷ و تعداد ژن‌ها ۳ عدد انتخاب شد. گام چهارم انتخاب تابع پیوندی است که در این مطالعه عمل جمع برای برای ایجاد پیوند بین زیرشاخه‌ها مورد استفاده قرار گرفت. در نهایت، در گام پنجم عملگرهای ژنتیک و نرخ هریک از آنها انتخاب می‌شود.

شبکه عصبی مصنوعی

یک ساختار نوعی شبکه عصبی معمولاً از لایه ورودی، لایه میانی (مخفی) و لایه خروجی تشکیل شده است که در شکل ۲ نشان داده شده است. لایه ورودی یک لایه انتقال دهنده و وسیله‌ای برای تهیه کردن داده‌ها است. آخرین لایه یا لایه‌های خروجی شامل مقادیر پیش‌بینی شده توسط شبکه می‌باشد و بنابراین خروجی الگو را معرفی می‌کند و لایه‌های میانی و مخفی که از گره‌های پردازشگر تشکیل شده‌اند، محل پردازش داده‌ها است. تعداد لایه‌های مخفی و تعداد گره‌ها در هر لایه مخفی به طور معمول با روش آزمون و خطا مشخص می‌شوند. گره‌های لایه‌های مجاور در شبکه به طور کامل در ارتباط هستند. ورودی‌های هر گره ممکن است از متغیرهای ورودی یا از خروجی گره‌های دیگر حاصل شود. هر گره دارای یک تابع عملگر است که به آن تابع تبدیل‌گر نیز گفته می‌شود.



شکل ۲- یک شبکه متداول سه لایه که نوعی شبکه پیشرو نیز می‌باشد.

جهت برازش و پیش‌بینی جریان روزانه ایستگاه ارمند، آزمون نکویی برازش انجام شد. جدول ۱ نتایج آزمون نکویی برازش سری روزانه جریان را با استفاده از روش پورت مانتو نشان می‌دهد. با توجه به نتایج این جدول می‌توان نتیجه گرفت که مقدار آماره پورت مانتو (Q) برای الگوهای انتخاب شده، کمتر از مقدار جدول کای اسکور در سطح ۵ درصد است. بنابراین، نتایج حاصل از این روش صحت و کفایت الگوهای برازشی ARMA(2,8) و BL(2,8,1,1) را تایید می‌نماید.

با توجه به تایید الگوها، پارامترهای هر دو الگو محاسبه شده و از آن‌ها برای تولید آمار از سال آبی ۱۳۸۶ تا ۱۳۹۰ (به مدت ۵ سال) استفاده شد. شکل ۳ نمودار پراکندگی مقادیر مشاهده‌ای در مقابل مقادیر شبیه‌سازی شده، حاصل از الگوهای ARMA(2,8) و BL(2,8,1,1) را برای سری روزانه جریان رودخانه کارون در ایستگاه ارمند نشان می‌دهد. نتایج ارزیابی عملکرد الگوهای بسط داده شده در جدول ۲ ذکر شده است. با توجه به این جدول مشاهده می‌شود که الگوی ARMA(2,8) نتوانسته به خوبی دبی جریان رودخانه کارون را الگوبندی کند، که این امر به دلیل خاصیت غیرخطی مقیاس روزانه می‌باشد. دقت بالای الگوی غیرخطی ناشی از تطابق ماهیت جریان رودخانه با الگوی منتخب می‌باشد. به طوری که ضریب همبستگی در الگو BL(2,8,1,1) معادل ۰/۹۱۳ محاسبه گردید.

نتایج الگوبندی جریان روزانه ایستگاه ارمند با استفاده از الگوهای هوش مصنوعی

الگوهای مورد استفاده در مقیاس زمانی روزانه در جدول ۳ ذکر شده و جدول ۴ مشخصات برنامه‌ریزی بیان ژن به کار رفته را برای سری روزانه نشان می‌دهد. جدول ۵ مقایسه شاخص‌های آماری مربوط به نتایج حاصل از کاربرد روش برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی در الگوبندی جریان روزانه رودخانه را برای الگوهای مختلف ورودی نشان می‌دهد.

آموزشی از آنها یاد می‌شود) به منظور آموزش ماهیت سازوکار حاکم بر پدیده نه تنها سبب پیچیدگی الگو و افزایش حافظه درگیر خواهد شد، بلکه سبب کاهش دقت الگو نیز می‌شود. لذا در الگوسازی جریان رودخانه‌ها نیز باید سعی نمود موثرترین داده‌های مشاهداتی را به عنوان داده‌های آموزشی انتخاب کرد.

معیارهای ارزیابی الگوها

نمایه‌های ضریب همبستگی و جذر میانگین مربعات خطا برای ارزیابی الگوهای مورد مطالعه در این پژوهش استفاده شده است (روابط ۱۳ و ۱۴). در صورتی که مقدار ضریب همبستگی بالا و خطا کم باشد، می‌توان نتایج دقیق‌تر و قابل اعتمادتری به دست آورد.

$$R = \sqrt{1 - \frac{\sum_{i=1}^n (Q_i - \bar{Q})^2}{\sum_{i=1}^n (Q_i - \bar{Q})^2}} \quad (11)$$

$$RMSE = \sqrt{\frac{\sum_{i=1}^n (Q_i - \bar{Q}_i)^2}{n}} \quad (12)$$

که در روابط بالا، Q_i مقدار مشاهده شده دبی روزانه رودخانه در گام زمانی i ام، \bar{Q}_i مقدار دبی تخمین زده شده در همان زمان، n تعداد داده‌ها و \bar{Q} میانگین مقادیر دبی مشاهداتی می‌باشد.

نتایج و بحث

نتایج الگوبندی غیرخطی سری روزانه رودخانه کارون

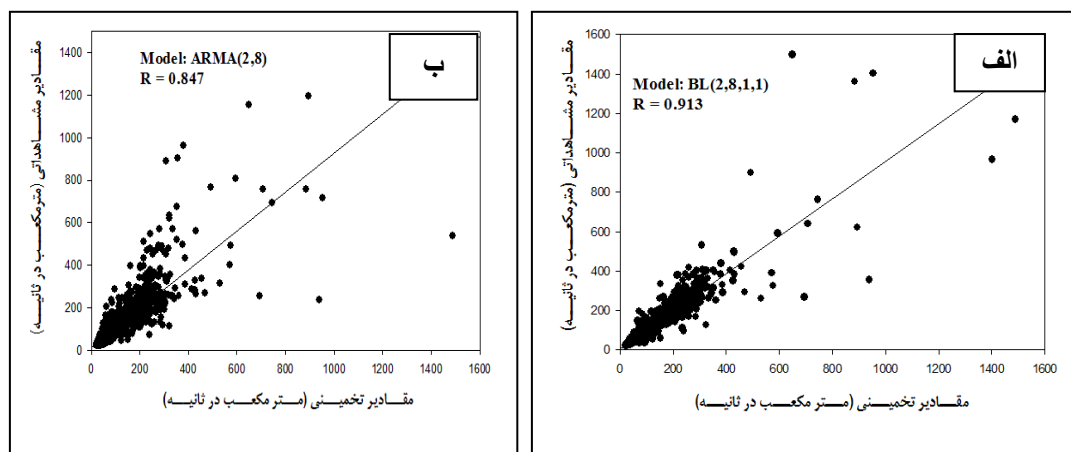
در این قسمت از هر دو الگوی خطی و غیرخطی استفاده شده و نتایج حاصل مورد مقایسه قرار گرفته است. برای سری روزانه جریان رودخانه کارون، الگوهای ARMA(2,8) و BL(2,8,1,1) با داشتن کمترین مقدار آکائیکه اصلاح شده معادل ۱۳۳۴/۱۰ و ۱۳۵۸/۱۲ به عنوان مناسب‌ترین الگوی خطی و غیرخطی انتخاب شدند. همچنین قابل ذکر است که قبل از استفاده از الگوهای منتخب

جدول ۱- نتایج آزمون پورت مانتو الگوهای خطی و غیرخطی برازشی جریان روزانه رودخانه کارون.

الگو	تعداد داده	تاخیر	درجه آزادی	سطح معنی داری	آزمون Q	آماره جدول	نتیجه آزمون
ARMA(2,8)	۸۹۳۵	۲۰۰	۱۹۰	۵ درصد	۱۸۳/۴۰	۲۳۴	پذیرش فرض صفر
BL(2,8,1,1)	۸۹۳۵	۲۰۰	۱۸۸	۵ درصد	۱۸۳/۱۹	۲۳۴	پذیرش فرض صفر

جدول ۲- تحلیل آماری دقت الگوبندی خطی و غیر خطی جریان روزانه رودخانه کارون

الگوی برازشی	ضریب همبستگی	RMSE (m ³ /s)
ARMA(2,8)	۰/۸۴۷	۵۸/۷۲
BL(2,8,1,1)	۰/۹۱۲	۴۲/۹۳



شکل ۳- مقایسه نتایج الگوبندی غیرخطی سری روزانه رودخانه کارون (الف: مدل $BL(2,8,1,1)$ ، ب: مدل $ARMA(2,8)$)

رودخانه کارون نشان می‌دهد. شکل‌های ۴ و ۵ به ترتیب نمودار مقادیر مشاهده‌ای و مقادیر شبیه‌سازی شده حاصل از بهترین الگوی شبکه عصبی و برنامه‌ریزی بیان ژن برای سری روزانه جریان رودخانه کارون در مرحله شبیه‌سازی (دوره آماری $(1386-1385)$ تا $(1390-1389)$) را نشان می‌دهند. با توجه به این شکل‌ها می‌توان نتیجه گرفت که برنامه‌ریزی بیان ژن در بهترین حالت خود به خوبی توانسته دبی جریان روزانه رودخانه مورد مطالعه را پیش‌بینی کند.

با توجه به نتایج این جدول مشاهده می‌شود الگوبندی با روش برنامه‌ریزی بیان ژن که در آن الگوی با دبی سه روز قبل کمترین خطا و بیشترین دقت را داشته است، به‌عنوان بهترین الگوی ساخته شده برای برنامه‌ریزی بیان ژن انتخاب گردید. همچنین در مقایسه با الگوهای خطی و غیرخطی سری‌زمانی روش برنامه‌ریزی ژنتیک نتیجه بهتری ارائه کرده است. رابطه ۱۳ شکل ریاضی به‌دست آمده از برنامه‌ریزی بیان ژن را در الگوی سوم، برای سری روزانه جریان

جدول ۳- الگوهای ورودی مورد استفاده در مقیاس زمانی روزانه رودخانه کارون

ردیف	الگوی ورودی روزانه
۱	$Q(t)=f\{Q(t-1)\}$
۲	$Q(t)=f\{Q(t-1), Q(t-2)\}$
۳	$Q(t)=f\{Q(t-1), Q(t-2), Q(t-3)\}$
۴	$Q(t)=f\{Q(t-1), Q(t-2), Q(t-3), Q(t-4)\}$

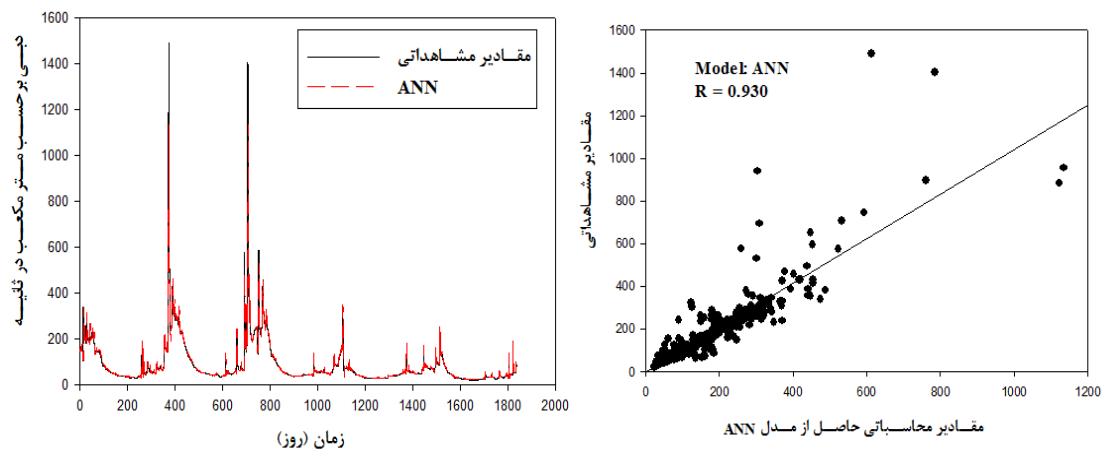
جدول ۴- مشخصات برنامه‌ریزی بیان ژن به‌کار گرفته شده

ردیف	معیار مورد نظر	مقدار عددی
۱	تولید بدون بهبود	۱۵۰
۲	تولید از زمان شروع	۲۰۰
۳	حداکثر مقدار اجرا	۱۵۰
۴	اندازه حداکثر برنامه	۵۱۲

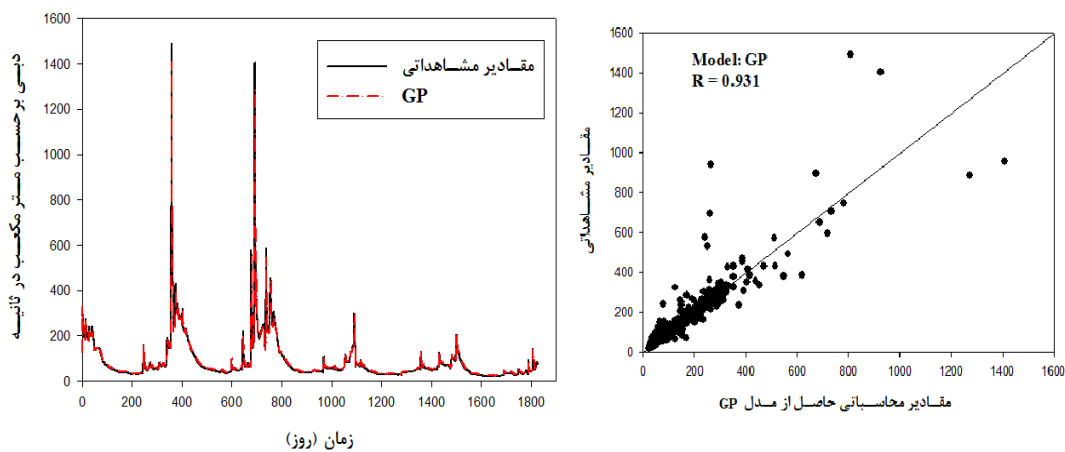
جدول ۵- تحلیل‌های آماری الگوهای هوش مصنوعی برای الگوهای مختلف ورودی جریان روزانه

شماره الگو	مرحله صحت‌سنجی ANN		مرحله صحت‌سنجی GEP	
	ضریب همبستگی	RMSE (m^3/s)	ضریب همبستگی	RMSE (m^3/s)
۱	۰/۹۱۰	۴۷/۱۷	۰/۹۱۷	۳۹/۳۵
۲	۰/۹۲۱	۳۹/۶۰	۰/۹۳۱	۳۸/۲۵
۳	۰/۹۲۸	۳۸/۰۸	۰/۹۳۱	۳۶/۵۹
۴	۰/۹۱۹	۴۰/۰۵	۰/۸۹۶	۴۲/۰۶

$$Q_t = Q_{t-1}^{0.5} + \frac{Q_{t-3} - 8.92}{\ln(Q_{t-3} \times Q_{t-2})} + Q_{t-1} - \frac{Q_{t-2}}{5.65} \quad (13)$$



شکل ۴- نتایج حاصل از شبیه‌سازی دبی روزانه رودخانه کارون با استفاده از شبکه عصبی مصنوعی



شکل ۵- نتایج حاصل از شبیه‌سازی دبی روزانه رودخانه کارون با استفاده از برنامه‌ریزی بیان ژن

هیدرولوژی و مدیریت منابع آب حوضه آبریز می‌باشد. مطالعه این پدیده هیدرولوژیکی دارای کاربردهای فراوانی از جمله در زمینه پیش-بینی سیلاب است. از طرفی به‌نظر می‌رسد فرآیند جریان رودخانه در مقطع زمانی کوتاه‌مدت به‌ویژه روزانه غیرخطی باشد. منبع اصلی غیرخطی بودن فرآیند جریان‌های کوتاه مدت روزانه از غیرخطی بودن فرآیندهای کوتاه مدت اقلیمی نظیر بارش و درجه حرارت منشعب شده و غیرخطی بودن واکنش بارش رواناب نیز باعث پیچیده‌تر شدن فرآیند رودخانه می‌گردد. یکی از کاربردهای این پژوهش معرفی و استفاده از الگوی غیرخطی دوخطی در منابع آب می‌باشد. این الگو با توجه به نتایج بدست آمده در این پژوهش با سری زمانی جریان رودخانه هم‌خوانی داشته و باعث بهبود نتایج نسبت به الگوهای خطی می‌شود. استفاده از این الگو به‌دلیل شباهت زیاد به ساختار الگوهای سری‌زمانی راجع ARMA و سادگی ساختار آن نسبت به سایر

با توجه به نتایج ذکر شده در جدول ۵ می‌توان با توجه به معیارهای ارزیابی مورد استفاده روش برنامه‌ریزی بیان ژن را مناسب‌تر از روش شبکه عصبی در الگوبندی جریان روزانه رودخانه کارون محسوب کرد. نتایج فوق با نتایج پژوهش ناوه و همکاران (۲۵) مطابقت دارد. همچنین مطابق با نتایج حاصل از پژوهش گون و همکاران (۱۸) روش برنامه‌ریزی بیان ژن یک روش مناسب در الگوبندی جریان روزانه رودخانه می‌باشد. مطابق با نتایج این پژوهش و پژوهش‌های فرودانم و همکاران (۱۱) و داندنمه‌ر و طباطبایی (۸)، روش برنامه‌ریزی بیان ژن از دقت بیشتری نسبت به شبکه عصبی مصنوعی برخوردار است.

نتیجه‌گیری

تعیین جریان رودخانه‌ها یکی از مولفه‌های مهم و تاثیرگذار در

ارزیابی، می‌توان در مجموع روش برنامه‌ریزی بیان ژن را به‌عنوان یک روش صریح و دقیق برای پیش‌بینی جریان رودخانه‌ها پیشنهاد کرد. عملکرد روش برنامه‌ریزی بیان ژن بهتر از شبکه عصبی و سری زمانی غیرخطی بوده است. پیشنهاد می‌شود در پژوهش‌های آینده از دیگر الگوهای هوش مصنوعی از جمله الگوی سیستم استنتاجی فازی-عصبی نیز در الگوبندی جریان رودخانه در مقیاس‌های زمانی روزانه (ماهانه) و سایر پارامترهای هیدرولوژیکی استفاده شود.

الگوهای غیرخطی سری زمانی می‌تواند در هیدرولوژی و منابع آب به سهولت مورد استفاده قرار گیرد. نتایج این مطالعه نیز نشان داد که الگوی دوخطی با دقت بالا و خطای کمتری نسبت به الگوی خطی توانسته جریان روزانه رودخانه کارون را در محل ایستگاه ارمند الگوبندی نماید.

در این پژوهش علاوه بر الگوی دوخطی از برنامه‌ریزی بیان ژن و شبکه عصبی مصنوعی نیز برای الگوبندی جریان روزانه رودخانه کارون در محل ایستگاه ارمند استفاده گردید. با توجه به معیارهای

منابع

- 1- Ainkaran P. 2004. Analysis of some linear and nonlinear time series models. A thesis submitted in fulfillment of the requirements for the degree of Master of Science, School of Mathematics and Statistics, University of Sydney.
- 2- Akaike H. 1974. A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control, AC-19, 716-723.
- 3- Alvisi S., Mascellani G., Franchini M., and Bardossy A. 2005. Water level forecasting through fuzzy logic and artificial neural network approaches. J Hydrol Earth Syst, Sci. 2: 1107-1145.
- 4- Aytak A., and Asce M. 2008. An application of artificial intelligence for rainfall runoff modeling. J Hydrol Earth Syst, Sci. 117(2): 145-155.
- 5- Baareh A.K.M., Sheta A.F., and Khnaifes K.A. 2006. Forecasting river flow in the USA : A comparison between auto- regression and neural network non-parametric models. J Comput Sci, 2(10): 775-780.
- 6- Caiado J. 2007. Forecasting water consumption in Spain using univariate time series models. Munich Personal RePEc Archive, No: 6610.
- 7- Dai Y. and Billard L. 2003. Maximum likelihood estimation in space time bilinea models. Journal of Time Series Analysis, 24(1): 25-44.
- 8- Dannademehr A. and Tabatabai M.R.M. 2010. Prediction of Daily discharge trend of river flow based on genetic programming. J. of Water and soil. 24(2): 325-333. (In Persian).
- 9- Dogan E., Isik S., Toluk T., and Sandalci M. 2007. Daily streamflow forecasting using artificial neural networks. Pp. 448-459. International Congress River Flood Management. Ankara, Turkey.
- 10- Fan J., and Yao Q. 2003. Nonlinear time series, nonparametric and parametric methods. Springer-Verlag, New York, Inc.
- 11- Farbodfam N., Ghorbani M.A., and Aalami M.T. 2009. Forecasting river flow using genetic programming (Case study: Lighwan watershed). J. Agric. Sci. 19(1): 4. 107-123. (In Persian).
- 12- Ferreira C. 2001. Gene expression programming: A new adaptive algorithm for solving problems. Complex Systems, 13(2): 87-129.
- 13- Firat M. 2007. Artificial intelligence techniques for river flow forecasting in the Seyhan river catchment, Turkey. J Hydrol Earth Syst, Sci. 4: 1369-1406.
- 14- Ghorbani M.A., Shiri J., and Kazemi H. 2010. Estimation of Maximum, Mean and Minimum Air Temperature in Tabriz City Using Artificial Intelligent Methods. J. Agric, Sci. 20(3): 4. 87-104. (In Persian).
- 15- Ghorbani M.A., Kisi O., and Aalinezhad M. 2010. A probe into the chaotic nature of daily streamflow time series by correlation dimension and largest Lyapunov methods. Applied Mathematical Modelling, 34: 4050-4057.
- 16- Govindaraju R.S. 2000. Artificial neural network in hydrology. Journal of hydrologic Engineering, 5(2): 115-123.
- 17- Granger C.W.J. and Andersen A.P. 1978. An Introduction to Bilinear Time Series Models. Vandenhoeek and Ruprecht: Göttingen.
- 18- Guven A. 2009. Linear genetic programming for time-series modeling of daily flow rate. J Earth Syst, Sci. 118(2): 157-173.
- 19- Khalili K., Fakheri Fard A., Dinpashoh Y., and Ghorbani M.A. 2011. Nonlinearity and testing for stream flow processes by BDS test (Case study: Shaharchai river). J. Agric. Sci. 22(2): 25-37. (In Persian).
- 20- Khalili K., Ahmadi F., Abghari H., and Besharat S. 2013. Stationary and nonlinearity analysis of river flow at different time scales (Case study: Barandouz-Chai River). 9th International River Engineering Conference, Ahwaz, Iran.
- 21- Khu S.T., Liong S.Y., Babovic V., Madsen H., and Muttill N. 2001. Genetic programming and its application in real- time runoff forming. Journal of American Water Resources Association, 37(2): 439-451.
- 22- Kisi O. 2007. Streamflow Forecasting Using Different Artificial Neural Network Algorithms. Journal of Hydrologic Engineering, 12(5): 533-539.

- 23- Lifshits M.A. 2006. Invariance principle in a bilinear model with weak nonlinearity. *Journal of Mathematical Sciences*, 137(1), 4541-4545.
- 24- Lohani A.K., Kumar R. and Singh R.D. 2012. Hydrological time series modeling: A comparison between adaptive neuro-fuzzy, neural network and autoregressive techniques. *J. Hydrology*, 442-443, 23-35.
- 25- Naveh H., Khalili K., Alami M.T., and Behmanesh J. 2012. Forecasting River flow By Bilinear Nonlinear Time Series Model (Case Study: Barandoz-Chay & Shahar-Chai Rivers). *J. of Water and soil*, 26(5): 1299-1307. (In Persian).
- 26- Salas J.D. 1993. Analysis and Modeling of Hydrological Time Series. In: *Handbook of Hydrology*, edited by David R. Maidment, McGraw-Hill, New York, 19(72).
- 27- Subba Rao T., and Gabr M. 1984. *An Introduction to Bispectral Analysis and Bilinear Time Series Models. Lecture Notes in Statistics*, 24, Springer-Verlag: New York.
- 28- Tsay R.S. 2002. *Analysis of financial time series*. University of Chicago, A Wiley – Interscience Publication, John Wiley f Sons; Inc.
- 29- Tsonis A.A. 2001. Probing the linearity and nonlinearity in the transitions of the atmospheric circulation. *Nonlinear Processes Geophysics*, 8: 341-345.
- 30- Wang W., Van G., Pieter H.A.J.M., Vrijling J.K., and Ma J. 2005. Testing and modelling autoregressive conditional heteroskedasticity of stream flow processes. *Nonlin. Processes Geophys.*, 12: 55-66, doi:10.5194/npg-12-55.
- 31- Wang W., Vrijling J.K., Pieter H.A.J.M., Van G., and Ma J. 2006. Testing for nonlinearity stream flow processes at different timescales. *J. Hydrology*, 322(1): 247-268.



Comparison of the Gen Expression Programming, Nonlinear Time Series and Artificial Neural Network in Estimating the River Daily Flow (Case Study: The Karun River)

R. Zamani^{1*} - F. Ahmadi² - F. Radmanesh³

Received:21-02-2014

Accepted:01-09-2014

Abstract

Today, the daily flow forecasting of rivers is an important issue in hydrology and water resources and thus can be used the results of daily river flow modeling in water resources management, droughts and floods monitoring. In this study, due to the importance of this issue, using nonlinear time series models and artificial intelligence (Artificial Neural Network and Gen Expression Programming), the daily flow modeling has been at the time interval (1981-2012) in the Armand hydrometric station on the Karun River. Armand station upstream basin is one of the most basins in the North Karun basin and includes four sub basins (Vanak, Middle Karun, Beheshtabad and Kohrang). The results of this study shown that artificial intelligence models have superior than nonlinear time series in flow daily simulation in the Karun River. As well as, modeling and comparison of artificial intelligence models showed that the Gen Expression Programming have evaluation criteria better than artificial neural network.

Keywords: Modeling, Artificial Intelligence, Daily discharge, Karun River

1, 2,3- PhD Students and Associate Professor, Department of Water Resources Engineering, Faculty of Water Science, Shahid Chamran University of Ahvaz, Iran, Respectively
(* - Corresponding Author Email: reza.zamani65@yahoo.com)