نشریه علمی-پژوهشی مهندسی و مدیریت آبخیز
جلد ۱۳، شماره ۳، ۱۴۰۰، صفحات ۶۱۴-۶۲۴

Journal of Watershed Engineering and Management
Volume 13, Issue 3, 2021, Pages 614-624
DOI: 10.22092/ijwmse.2021.126342.1656

تخمین جریان روزانه رودخانه با استفاده از مدل‌های هوشمند، مطالعه موردی: رودخانه مهاباد

عباس عباسی*[۱]، کیوان خلیلی[۲]، جواد بهمنش[۳] و اکبر شیرزاد[۴]

[۱] دانش‌آموخته دکتری مهندسی منابع آب، دانشکده کشاورزی، دانشگاه ارومیه، [۲] استادیار، دانشکده کشاورزی، دانشگاه ارومیه، [۳] استاد، دانشکده کشاورزی، دانشگاه ارومیه و [۴] دانشیار، دانشکده محیط زیست، دانشگاه صنعتی ارومیه

تاریخ دریافت: ۱۳۹۹/۰۲/۲۹ تاریخ پذیرش: ۱۳۹۹/۱۲/۲۵

چکیده

تخمین صحیح و دقیق جریان رودخانه می‌تواند نقش مهمی در کاهش اثرات ناشی از خسارات سیلاب ایفا کند. در این تحقیق، از مدل برنامه‌ریزی بیان‌ژن (GEP) و شبکه بیزین (BN) برای پیش‌بینی جریان روزانه رودخانه مهاباد واقع در حوزه آبخیز دریاچه ارومیه استفاده شد. بر این اساس، از چهار الگوی ورودی با تأخیرهای یک تا چهار روزه برای پیش بینی مقادیر جریان روزانه در زمان t+1 در یک دوره ۲۳ ساله استفاده و از ۷۵ درصد داده‌ها به‌منظور آموزش مدل‌ها و از ۲۵ درصد باقی‌مانده برای مرحله آزمون استفاده شد. نتایج نشان داد که الگوی برتر در هر دو روش، مدل با مقادیر ورودی تا سه گام زمانی تأخیر می‌باشد. همچنین، بر اساس سه شاخص ارزیابی ضریب همبستگی (R)، مجذور میانگین مربعات خطا (RMSE) و ضریب نش-ساتکلیف (E) در مرحله آزمون، روش برنامه‌ریزی بیان ژن با آماره‌های ارزیابی R=۰/۹۰۲ ، (m^3s^{-1}) RMSE=۲/۷۱ و E=۰/۸۱۲ نسبت به روش شبکه بیزین با آماره‌های ارزیابی R=۰/۹۰۵، (m^3s^{-1}) RMSE=۲/۶۷۹ و E=۰/۸۱۷ دارای دقت بالاتری می‌باشد. در حالت کلی، هر دو روش دارای دقت قابل قبول و نسبتاً یکسان هستند، ولی به‌دلیل مدل‌سازی آسان‌تر روش شبکه بیزین این مدل می‌تواند به‌عنوان یک روش کارآمد در پیش‌بینی جریان رودخانه‌ها مورد استفاده قرار گیرد.

واژه‌های کلیدی: برنامه‌ریزی بیان ژن، پیش‌بینی، دریاچه ارومیه، سیل، شبکه بیزین

مقدمه

پیش‌بینی جریان از جمله مسائل مهم در مدیریت عرضه و تقاضای آب محسوب می‌شود و امکان برنامه‌ریزی برای شرایط آتی را فراهم می‌کند که این امر به‌خصوص در شرایط بحرانی از قبیل وقوع سیلاب و خشکسالی می‌تواند نقش بسزایی در کاهش اثرات و آمادگی لازم برای رویارویی با آن‌ها ایفا کند. روش‌های مدل‌سازی جریان رودخانه‌ها در حالت کلی به دو دسته مدل‌های مفهومی یا فیزیکی و مدل‌های مبتنی بر داده یا آمار (یا روش‌های هوش مصنوعی) تقسیم‌بندی می‌شوند (Zhang و همکاران، ۲۰۱۶). اساس بیشتر روش‌های پیش‌بینی، نوعی شبیه‌سازی وضعیت موجود سامانه است که در اصطلاح به آن مدل سازی گفته می‌شود. مدل‌های مفهومی سعی می‌کنند که فرایندهای هیدرولوژیک همانند دبی رودخانه را به شیوه‌ای فیزیکی و واقع‌بینانه و از طریق قوانین فیزیکی مدل‌سازی کنند (Awchi، ۲۰۱۴). مدل‌های فیزیکی پیچیده بوده، به ابزارهای ریاضی پیچیده، مقدار قابل

* مسئول مکاتبات: abasi.abas2010@gmail.com

توجهی از داده‌ها برای واسنجی مدل‌ها و درجاتی از تخصص و تجربه کار با این مدل‌ها نیاز دارند (He و همکاران، ۲۰۱۴). با در نظر گرفتن مسائل و مشکلات مربوط به استفاده از مدل‌های مفهومی در مدل‌سازی دبی جریان، پژوهشگران به استفاده از مدل‌های آماری یا مبتنی بر داده روی آورده‌اند. طی سال‌های اخیر، مدل‌های سری زمانی به‌عنوان یکی از مدل‌های آماری، کاربرد گسترده‌ای در پیش‌بینی جریان رودخانه‌ها پیدا کرده‌اند. غالباً در روش‌های مرسوم برای پیش‌بینی سری‌های زمانی متغیرهای هیدرولوژیک، از فرض برقراری رابطه خطی بین متغیرها استفاده می‌شود. این روش‌ها، توسعه‌یافته الگوهای Box و همکاران (۲۰۱۱) هستند. این در حالی است که بسیاری از فرایندهای مربوط به سامانه‌های طبیعی همانند جریان رودخانه‌ها نسبت به زمان غیرخطی هستند و لذا، استفاده از انواع الگوهای خطی برای مدل‌سازی آن‌ها از دقت لازم برخوردار نیست. بنابراین، با توجه به اهمیت پیش‌بینی جریان در رودخانه‌ها، ارائه روشی مناسب در این خصوص بسیار مهم و حائز اهمیت می‌باشد. استفاده از روش‌های هوش مصنوعی در سالیان اخیر مورد توجه بسیاری از محققین منابع آب و هیدرولوژی برای مدل‌سازی و پیش‌بینی متغیرهای هیدرولوژیک همانند دبی جریان در رودخانه‌ها قرار گرفته است.

شبکه‌های بیزین[1] برای نشان دادن اطلاعات در یک حوضه دارای عدم قطعیت به‌کار می‌روند. شبکه‌های بیزین کاربردهای بسیاری در علوم مختلف داشته است که در سال‌های اخیر در مدیریت منابع آب و محیط‌زیست نیز کاربردهای موفقی از آن گزارش شده است. از جمله مطالعات انجام گرفته در رابطه با شبکه‌های بیزین می‌توان به پژوهش Brandt و Henriksen (۲۰۰۳) اشاره کرد که برای حفاظت و مدیریت کمی و کیفی آب شرب استحصالی از منابع آب زیرزمینی در دانمارک از شبکه بیزین استفاده کردند. نتایج این پژوهش، نشان داد که کاربرد شبکه‌های بیزین به‌عنوان یک ابزار تصمیم‌گیری در مدیریت منابع آب زیرزمینی موجود بسیار مؤثر می‌باشد. Reggiani و Weerts (۲۰۰۸) برای پیش‌بینی تراز سطح آب در رودخانه راین از مدل BN استفاده کردند. در این پژوهش، جهت برآورد ارزش اقتصادی اعلام هشدار سیلاب از یک تابع سود-هزینه استفاده کردند که در آن خسارت ناشی از هر تراز سطح آب پیش‌بینی شده تعیین شد. نتایج این پژوهش، نشان داد که در شرایط عدم قطعیت با استفاده از روش BN می‌توان پیش‌بینی جریان را با دقت مناسب پیش‌بینی نمود. Ahmadi و همکاران (۲۰۱۶) در تحقیقی جهت برآورد میزان آبدهی روزانه رودخانه باراندوزچای از دو روش شبکه‌بیزین و برنامه‌ریزی ژنتیک[2] (GP) استفاده کردند که نتایج این تحقیق نشان داد که روش برنامه‌ریزی ژنتیک نسبت به روش BN از دقت بالاتری برخوردار است و همچنین، این روش در تخمین دبی‌های کمینه و متوسط به مراتب بهتر از روش شبکه‌های بیزین عمل می‌کند. Ghorbani و Dehghani (۲۰۱۶) برای برآورد مواد جامد محلول در رودخانه بالخلوچای در استان اردبیل از روش‌های شبکه عصبی بیزین، شبکه عصبی مصنوعی و برنامه‌ریزی بیان ژن[3] استفاده کردند. نتایج این پژوهش، نشان داد که روش شبکه عصبی بیزین نسبت به دو روش دیگر از دقت بالاتری برخوردار است. Rezaei و همکاران (۲۰۱۹) نیز برای پیش‌بینی مواد جامد محلول در آب رودخانه کشکان از روش‌های ماشین بردار پشتیبان (SVM)، برنامه‌ریزی بیان ژن و شبکه بیزین استفاده کردند که نتایج این پژوهش، نشان داد که هر سه مدل دارای دقت قابل قبول در پیش‌بینی کیفیت آب هستند.

برنامه‌ریزی بیان ژن که در ادامه سیر تکاملی مدل‌های هوشمند به‌وجود آمده جزء روش‌های الگوریتم گردشی محسوب می‌شود که مبنای تمامی آن‌ها بر اساس نظریه تکامل داروین[4] استوار است. امروزه GEP در زمینه مهندسی آب به‌عنوان ابزاری مناسب برای مدل‌سازی مسائل مربوط به تعیین ساختار پدیده‌ها مطرح شده است. Dorado و همکاران (۲۰۰۳) فرایند بارش-رواناب را با هدف طراحی یک سامانه هشدار زمان واقعی سیلاب در حوضه‌های

[1] Bayesian networks

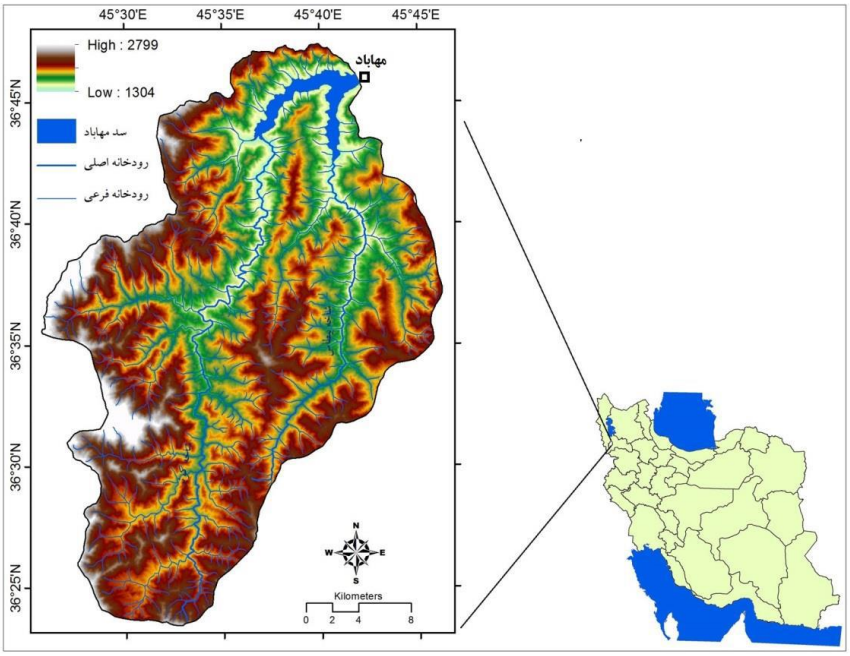
[2] Genetic Programming
[3] Gene Expression Programming
[4] Darwinian Evolution Theory

شهری با استفاده از روش GEP و شبکه عصبی مصنوعی مدل‌سازی کردند. نتایج این تحقیق، نشان داد که تجزیه و تحلیل حوضه‌های رودخانه‌ای با روش GEP و شبکه عصبی مصنوعی نسبت به روش‌های تصادفی از قبیل هیدروگراف واحد بهبود مدل‌سازی بارش–رواناب را در پی داشته است. به‌منظور پیش‌بینی جریان روزانه رودخانه شویل کیل در آمریکا، Guven (۲۰۰۹) از روش GEP استفاده کرده، نتایج آن را با دو الگو از روش شبکه عصبی مصنوعی مقایسه کرد. نتایج این پژوهش، نشان داد که با وجود این‌که روش GEP در برآورد میزان دبی از دقت بالاتری برخوردار است، اما روش شبکه عصبی مصنوعی نیز دارای دقت قابل قبولی است. Khu و همکاران (۲۰۰۱) نیز برای پیش‌بینی جریان رودخانه در گام زمانی ساعتی در یکی از حوزه‌های آبخیز فرانسه از روش GEP استفاده کردند و نتایج آن را با روش رگرسیون و فیلتر کالمن مورد مقایسه قرار دادند. مقایسه نتایج این پژوهش با مطالعه دیگری که از روش خودهمبستگی و فیلتر کالمن برای پیش‌بینی رواناب استفاده شده بود، حاکی از برتری روش GEP است. Ahmadi و همکاران (۲۰۱۶) نیز از روش برنامه‌ریزی ژنتیک و ماشین‌بردار پشتیبان به‌منظور پیش‌بینی جریان روزانه رودخانه باراندوزچای استفاده کردند که نتایج نشان داد، هر دو مدل دارای دقت قابل قبول در برآورد میزان دبی هستند و دقت روش GP از روش SVM اندکی بیشتر است. Terzi و Ergin (۲۰۱۴) دبی جریان ماهانه رودخانه کیزایرماق در ترکیه را با مدل‌های هوشمند GEP، ANN و ANFIS و نیز مدل سری زمانی AR طی دوره آماری ۱۹۵۵ تا ۱۹۹۵ برآورد کردند. در فرایند مدل‌سازی، مقادیر دبی ماهانه با تأخیرهای زمانی از یک تا سه ماه به‌عنوان ورودی به مدل‌ها وارد شدند. نتایج نشان داد که مدل AR که از مقادیر دبی دو ماه قبل برای تخمین دبی ماه جاری استفاده می‌کند، بیشترین دقت تخمین را داشته است. همچنین، در بین مدل‌های هوشمند، مدل‌های GEP و ANFIS بر مدل ANN برتری داشتند.

از آنجا که پیش‌بینی جریان رودخانه‌ها نقش مهمی در برنامه‌ریزی و مقابله با خطرات ناشی از سیلاب‌ها بر عهده دارد و با توجه به بحران آبی حاکم بر دریاچه ارومیه و اهمیت رودخانه مهاباد در تأمین نیاز زیست‌محیطی آن، هدف از مطالعه حاضر پیش‌بینی جریان روزانه این رودخانه با استفاده از مدلی است که بتواند از دقت و کارایی بالا برخوردار باشد. بدین‌منظور، در تحقیق حاضر سعی شده است با استفاده از دو روش شبکه بیزین و برنامه‌ریزی بیان ژن مقادیر جریان برای یک روز آینده برآورد شود و در ادامه دقت مدل‌های پیش‌بینی از طریق مقایسه‌ی مقادیر پیش‌بینی شده با مقادیر مشاهداتی و با استفاده از آماره‌های ارزیابی مورد مقایسه قرار می‌گیرد و مدل پیش‌بینی برتر معرفی شود.

مواد و روش‌ها

منطقه مورد پژوهش و داده‌های مورد استفاده: رودخانه مهاباد یکی از منابع مهم آب سطحی دشت مهاباد و یکی از رودخانه‌های مهم و پرآب حوزه آبخیز دریاچه ارومیه می‌باشد. حوزه آبخیز رودخانه مهاباد حدود ۸۵۰ کیلومتر مربع وسعت دارد. این حوضه در قسمت غربی به‌وسیله خط الرأس‌های ارتفاعات زاگرس از حوضه زاب کوچک و به‌وسیله ارتفاعات سپیامان از حوزه آبخیز گدار و به‌وسیله ارتفاعات شیخ حیدر، شاه-قل و مام سوار از حوزه آبخیز سیمینه‌رود جدا می‌شود. آب رودخانه در زمان ذوب برف‌ها و یا نزول شدید بارش طغیان کرده، حدود ۱۰ متر از بستر اصلی آن بالا می‌آید، لذا، به‌منظور جلوگیری از هدر رفتن این آب و ذخیره آن، سد مخزنی مهاباد در مسیر این رودخانه احداث شده است. طول رودخانه مهاباد حدود ۹۰ کیلومتر است. شکل ۱، موقعیت حوزه آبخیز رودخانه مهاباد را نشان می‌دهد. داده‌های مورد استفاده در این تحقیق، داده‌های دبی روزانه رودخانه مهاباد در ایستگاه هیدرومتری کوتر طی دوره آماری ۱۳۹۲–۱۳۶۹ می‌باشد که از شرکت آب منطقه‌ای آذربایجان غربی دریافت شده است و پس از بررسی صحت و سقم آن‌ها با استفاده از آزمون‌های آماری و تکمیل و تصحیح آن‌ها برای استفاده در مدل‌های پیش‌بینی با چهار گام زمانی تأخیر آماده‌سازی و به‌کار گرفته شدند.



شکل ۱- نقشه حوزه آبخیز رودخانه مهاباد

شبکه بیزین: شبکه‌های بیزین در واقع یک نوع خاص از مدل‌های گرافیکی احتمالاتی هستند که مجموعه‌ای از متغیرها و احتمالات مربوط به هر کدام را نشان می‌دهند. در حالت کلی، شبکه‌های بیزین برای حالاتی مفیدند که وضعیت فعلی سامانه به وضعیت قبلی آن بستگی دارد. ساختار BN در واقع یک نمایش گرافیکی از اثرات متقابل متغیرهایی است که بایستی مدل‌سازی شوند. در شبکه‌های بیزین، یک گراف متشکل از مجموعه گره‌ها و کمان‌های ارتباطی است. گره‌ها نیز نماینده متغیرهای تصادفی گسسته‌سازی شده هستند. به دو گره یا متغیر که به‌وسیله یک کمان ارتباطی به هم متصل شده باشند، یک لبه می‌گویند. برای هر دو متغیر که ارتباط سببی بین آن‌ها وجود دارد، لبه به‌صورت یک خط جهت‌دار است که متغیر تأثیرگذار را به متغیر تأثیرپذیر وصل می‌کند. توزیع احتمالاتی توأم یک مجموعه‌ای از متغیرها با فرض مستقل بودن آن‌ها از حاصل‌ضرب توزیع احتمالاتی شرطی آن‌ها از طریق رابطه (۱) به دست می‌آید (Neapolitan، ۲۰۰۳).

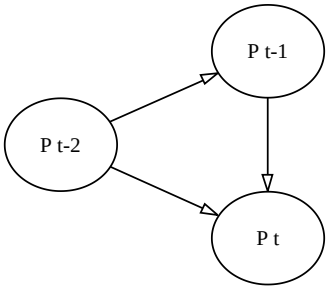
$$P(x_1, x_2, \dots, x_n) = \prod_{i=1}^{n} P(x_i|\pi(x_i)) \qquad (۱)$$

که در آن، x_i iامین مقدار متغیر تصادفی، $P(x_1, x_2, \dots, x_n)$ احتمال توأم مقادیر $(x_1, x_2, \dots, x_n)$ و $\pi(x_i)$ ارزش یا مقادیر کمی مجموعه علل x_i هستند. در این پژوهش، سه گام زیر برای مدل‌سازی دبی رودخانه با استفاده از شبکه‌های بیزین به شرح زیر انجام گرفت.

۱- ایجاد مدل مفهومی بر اساس ورودی‌های در نظر گرفته شده برای مدل‌سازی متغیر مستقل دبی: نمونه‌ای از شبکه‌های بیزین ایجاد شده برای مدل‌سازی دبی با استفاده از دو متغیر در شکل ۲ نشان داده شده است؛ به‌طوری‌که مقدار دبی روز آتی با استفاده از مقادیر دبی دو روز پیشین مدل و برآورد می‌شود.

۲- ایجاد ارتباطات مابین متغیرهای مدل مفهومی ایجاد شده با استفاده از کمان‌های ارتباطی: خطوط جهت‌دار از P_{t-1} و P_{t-2} به سمت P_t نمایانگر این نکته است که مقدار دبی روز آتی (P_t) به‌وسیله مقادیر دبی دو روز قبلی (P_{t-1} و P_{t-2}) تحت تأثیر قرار می‌گیرد.

۳- آموزش شبکه بیزین توسعه داده شده: پس از آموزش شبکه، مدل بیزین رابطه ریاضی ارائه می‌کند که این معادله برای به‌دست آوردن مقادیر پیش‌بینی شده به‌وسیله شبکه‌های بیزین مورد استفاده قرار می‌گیرد.



شکل ۲- مدل مفهمومی ایجاد شده برای مدل‌سازی دبی با دو گام زمانی تأخیر

برنامه‌ریزی بیان ژن (GEP): برنامه‌ریزی بیان ژن، تعمیم‌یافته الگوریتم ژنتیک است که برای اولین‌بار به‌وسیله Ferreira (۲۰۰۱) و بر اساس نظریه داروین پیشنهاد و معرفی شد. برنامه‌ریزی بیان ژن همانند الگوریتم ژنتیک و برنامه‌ریزی ژنتیک است که از جمعیتی از افراد استفاده کرده، آن‌ها را بر اساس برازندگی انتخاب می‌کند و تغییرات ژنتیکی را با استفاده از یک یا چند عملگر ژنتیکی اعمال می‌کند. تفاوت اساسی بین سه الگوریتم یاد شده مربوط به ماهیت افراد آن‌هاست. در الگوریتم ژنتیک افراد به‌صورت رشته‌های خطی با طول ثابت (کروموزوم‌ها) هستند. در حالی که در برنامه‌ریزی ژنتیک، افراد به‌صورت نهاده‌های غیرخطی با اندازه‌ها و اشکال متفاوت (درختان تجزیه) هستند. این در حالی است که در GEP، افراد در ابتدا به‌صورت رشته‌های خطی با طول ثابت کدگذاری می‌شوند (مشابه با الگوریتم ژنتیک) و سپس، به شکل نهاده‌های غیرخطی با اندازه‌ها و اشکال متفاوت (مشابه با برنامه‌ریزی ژنتیک) بیان می‌شوند. تولید جمعیت اولیه از راه‌حل‌ها، اولین مرحله در روش GEP است. این امر می‌تواند با استفاده از اطلاعات موجود یا پیشین درباره مسئله مورد بررسی و یا به‌صورت تصادفی انجام گیرد. کروموزوم‌ها به‌صورت بیان درختی نشان داده می‌شوند. سپس، برازش‌های صورت گرفته با استفاده از توابع برازش مورد ارزیابی قرار می‌گیرند. اگر یک کیفیت مطلوبی از راه‌حل مسئله یافت شود و یا این‌که نسل‌ها به تعداد معینی برسند، فرایند تکامل (تکرار) متوقف می‌شود و بهترین راه‌حل یافت شده گزارش می‌شود. با این حال، اگر شرایط توقف برنامه GEP حاصل نشود، بهترین راه‌حل یافت شده نگه داشته می‌شود که این امر به فرایند نخبه‌گزینی معروف است. در حقیقت، افراد شایسته شانس بیشتری برای تولید فرزندان دارند. پس از آن تغییرات ژنتیکی با استفاده از عملگرهای ژنتیکی مختلف اعمال می‌شود. این روند برای نسل‌های موجود تکرار می‌شود تا کیفیت رضایت‌بخشی از راه‌حل مسئله یافت شود. در روش GEP، پدیده‌های مختلف با استفاده از مجموعه‌ای از توابع و مجموعه‌ای از ترمینال‌ها برآورد و مدل‌سازی می‌شوند. مجموعه توابع شامل توابع اصلی حسابی و یا توابع تعریف شده به‌وسیله کاربر هستند که این توابع می‌توانند برای تفسیر مدل مناسب باشند. مجموعه ترمینال‌ها نیز از متغیرهای مستقل مسئله و مقادیر ثابت تشکیل شده‌اند. در این تحقیق، از نرم‌افزار GeneXproTools 4.0 به‌منظور اعمال روش GEP بر روی داده‌ها استفاده شده است. فرایند گام به گام برای مدل‌سازی متغیرهای در نظر گرفته شده در پژوهش حاضر با استفاده از مدل GEP به‌صورت زیر است.

نخستین گام، انتخاب تابع برازش است که می‌تواند بر اساس روابط متعددی صورت گیرد که در این پژوهش از تابع برازش ریشه میانگین مربعات خطا (RMSE) استفاده شده است. در گام دوم، مجموعه متغیرهای ورودی و مجموعه توابع به‌منظور تولید کروموزوم‌ها انتخاب و معرفی می‌شوند. در این پژوهش، مجموعه متغیرهای ورودی شامل متغیرهای مستقلی (دبی با تأخیرهای زمانی یک تا چهار روزه) هستند که به‌عنوان ورودی برای مدل‌سازی متغیر وابسته (دبی در یک روز آینده) مد نظر قرار گرفتند. همچنین، از چهار عملگر اصلی برای ایجاد ارتباط بین متغیرها شامل $\{+,-,\times,\div\}$ استفاده شد. در گام سوم ساختار و معماری کروموزوم‌ها انتخاب شدند. به‌طوری‌که اندازه طول هر رأس و تعداد ژن‌ها به‌ترتیب برابر هشت و سه مورد استفاده قرار گرفت. گام چهارم انتخاب تابع پیوندی است که تابع جمع برای ایجاد پیوند بین زیرشاخه‌ها استفاده شد. در گام پنجم نیز عملگرهای ژنتیک و نرخ هر یک از آن‌ها انتخاب شد.

در مدل‌های پیش‌بینی، تعیین بهترین الگوی ورودی برای آموزش مدل سبب افزایش عملکرد و دقت این مدل‌ها می‌شود. بدین‌منظور، در گام نخست باید بهترین متغیرهای ورودی به مدل پیش‌بینی برای

برآورد دقیق دبی مشخص شود. به‌منظور پیش‌بینی دبی، چهار الگوی مختلف ورودی مورد استفاده قرار گرفت. سپس، بر اساس معیارهای ارزیابی بهینه‌ترین الگو در پیش‌بینی دبی انتخاب شد. در الگوی ۱، برای پیش‌بینی دبی در روز آتی از سری زمانی دبی در روز جاری استفاده شد. در الگوی ۲، پیش‌بینی دبی در روز بعد بر اساس مقادیر دبی تا یک روز قبل صورت پذیرفت و به‌طور مشابه در الگوهای ۳ و ۴ پیش‌بینی بر اساس مقادیر شاخص تا دو و سه روز قبل انجام شد (روابط ۲ الی ۵).

$$Q_{(t+1)} = f(Q_{(t)} \quad (۲)$$

$$Q_{(t+1)} = f(Q_{(t)}, Q_{(t-1)} \quad (۳)$$

$$Q_{(t+1)} = f(Q_{(t)}, Q_{(t-1)}, Q_{(t-2)}) \quad (۴)$$

$$Q_{(t+1)} = f(Q_{(t)}, Q_{(t-1)}, Q_{(t-2)}, Q_{(t-3)}) \quad (۵)$$

به‌منظور پیش‌بینی جریان روزانه رودخانه مهاباد در این تحقیق، از ۷۵ درصد داده‌ها به‌منظور آموزش مدل‌های شبکه بیزین و برنامه‌ریزی بیان ژن و از ۲۵ درصد باقی‌مانده، برای آزمون آن‌ها استفاده شده است.

معیارهای ارزیابی مدل: در این تحقیق، به‌منظور ارزیابی مدل‌های مورد نظر از معیارهای ضریب همبستگی (R)، نش-ساتیکلف (E) و جذر میانگین مربعات خطا ($RMSE$) استفاده شد که روابط آن‌ها به‌ترتیب به‌صورت روابط (۶) الی (۸) است.

$$R = \left(1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i - \bar{Q}_i)^2}\right)^{0.5} \quad (۶)$$

$$E = 1 - \frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{\sum_{i=1}^{n}(Q_i - \bar{Q}_i)^2} \quad (۷)$$

$$RMSE = \left(\frac{\sum_{i=1}^{n}(Q_i - \hat{Q}_i)^2}{n}\right)^{0.5} \quad (۸)$$

که در آن‌ها، Q_i مقدار داده مشاهداتی در گام زمانی iام، $\hat{Q}_i$ مقدار داده پیش‌بینی شده در همان زمان، n تعداد داده‌ها و $\bar{Q}_i$ میانگین مقادیر مشاهداتی است. هرچه مقادیر R و E به یک نزدیک‌تر و RMSE به صفر نزدیک‌تر باشد، نتایج دقیق‌تر و قابل اعتمادتر می‌باشد.

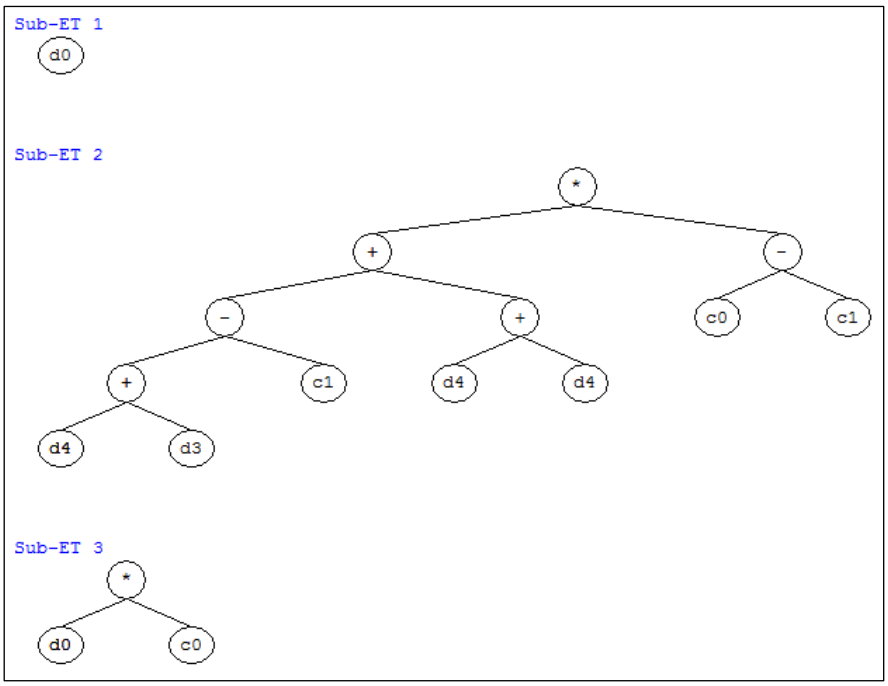
نتایج و بحث

نتایج مدل‌سازی به روش برنامه‌ریزی بیان ژن: در این مطالعه، توالی دبی در چهار تأخیر زمانی به‌عنوان الگوهای ورودی به مدل برای آموزش استفاده شده است. یکی دیگر از موارد مهم در GEP انتخاب عملگرهای ریاضی به‌منظور دستیابی به رابطه بین متغیر وابسته و متغیرهای ورودی مستقل می‌باشد که روابط اصلی جمع، تفریق، تقسیم و ضرب به‌عنوان عملگرهای اصلی انتخاب شده‌اند. مشخصات مدل GEP استفاده شده در این مطالعه در جدول ۱ ارائه شده است.

جدول ۱- مشخصات مدل GEP مورد استفاده

ردیف	معیار مورد نظر	مقدار عددی
۱	میزان جمعیت تولید شده	۱۰۰۰
۲	بهترین برازش	۱۳۷
۳	اندازه برنامه	۱۵
۴	تغییر در جمعیت تولید شده	۲۷

یکی از مزیت‌های مهم روش GEP ارائه رابطه ریاضی بین متغیر وابسته (دبی پیش‌بینی شده) و متغیرهای مستقل (دبی با تأخیر زمانی) است که این رابطه با استفاده از ساختار بیان درختی ارائه می‌شود. بیان درختی کمک می‌کند، در هر مرحله جمعیت اولیه به‌صورت ساختار خطی ساده بیان شود و تمام تغییرات تنها بر روی ساختارهای ساده انجام گیرد و در نتیجه نیازی به ساختارهای پیچیده برای گسترش در هر مرحله نخواهد بود. الگوهای مورد نظر در GEP به‌صورت ساختار درختی برای بهترین مدل پیش‌بینی در شکل ۳ ارائه شده است که در این شکل C_0 و C_1 ضرایب GEP و d_0 ، d_3 و d_4 به‌ترتیب معادل دبی بدون تأخیر، با سه و چهار روز تأخیر می‌باشد.



شکل ۳- ساختار درختی مدل GEP ژن برای بهترین مدل پیش‌بینی جریان روزانه رودخانه مهاباد

با استفاده از این ساختار درختی، رابطه ریاضی برای برآورد مقادیر جریان روزانه رودخانه مهاباد به‌صورت رابطه (۹) است.

$$Q_{t+1} = 0.853Q_t + 0.0142Q_{t-3} + 0.0426Q_{t-3} - 0.0133 \quad (۹)$$

نتایج مدل GEP برای دو مرحله آموزش و آزمون با متغیرهای مختلف ورودی در جدول ۲ ارائه شده است. بر اساس این جدول مشخص می‌شود که تا الگوی ۳ مقادیر آماره‌های ارزیابی به‌صورت مناسبی بهبود پیدا می‌کند و از الگوی ۳ به بعد تغییرات در بهبود آماره‌های ارزیابی ناچیز است، لذا، بر این اساس الگوی ۳ به‌عنوان مدل برتر انتخاب می‌شود. ضریب همبستگی برای الگوی برتر در مرحله آموزش و آزمون به‌ترتیب برابر ۰/۸۸ و ۰/۹۰۲ و RMSE نیز به‌ترتیب ۶/۰۶ و ۲/۷۱ متر مکعب بر ثانیه است که نشان از دقت قابل قبول این مدل دارد.

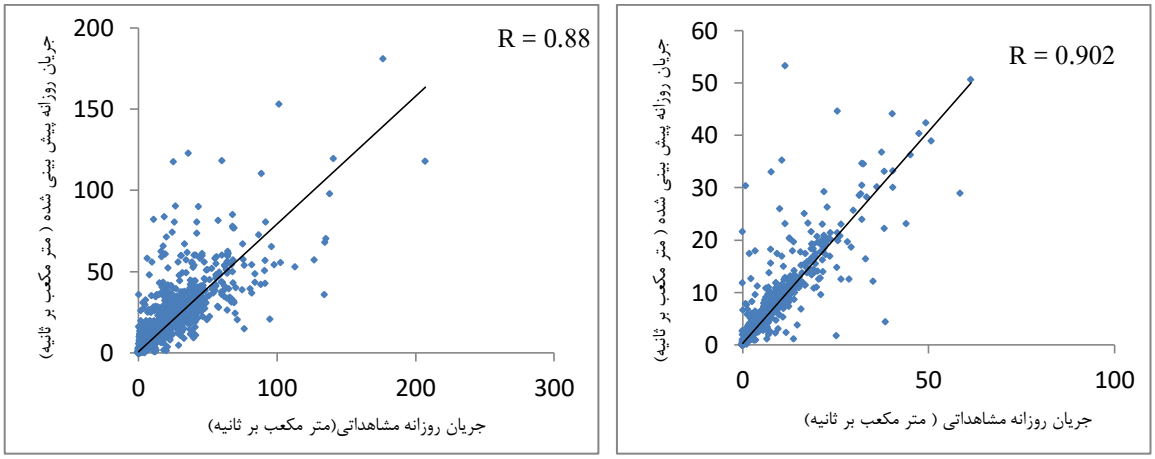
جدول ۲- معیارهای ارزیابی روش GEP در دو مرحله آموزش و آزمون

الگوی ورودی	دوره آموزش			دوره آزمون		
	R	RMSE (m^3s^{-1})	E	R	RMSE (m^3s^{-1})	E
۱	۰/۷۳	۸/۲۴	۰/۶۵۴	۰/۷۲۴	۸/۴۱	۰/۶۳۳
۲	۰/۸۱	۷/۰۷	۰/۷۱۸	۰/۷۹۲	۳/۷۱	۰/۷۴۸
۳	۰/۸۸	۶/۰۶	۰/۷۷۲	۰/۹۰۲	۲/۷۱	۰/۸۱۲
۴	۰/۸۸۵	۶/۰۱	۰/۷۸	۰/۹۰۶	۲/۶۷	۰/۸۱۵

پراکنش و سری زمانی مقادیر جریان روزانه مشاهداتی و جریان پیش‌بینی شده با الگوی برتر GEP برای دو مرحله آموزش و آزمون به‌ترتیب در شکل‌های ۴ و ۵ ارائه شده است که بر این اساس مشاهده می‌شود، در بیشتر موارد مقادیر پیش‌بینی شده با مقادیر مشاهداتی از روند یکسانی برخوردار هستند که نشان از دقت قابل قبول این مدل در پیش‌بینی روزانه مقادیر جریان رودخانه مهاباد دارد.

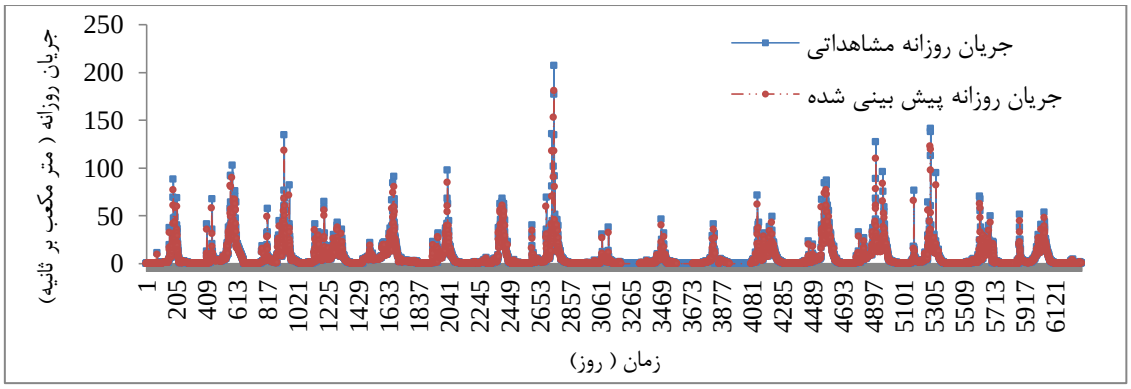
نتایج مدل‌سازی به روش شبکه بیزین: ساختار متغیرهای ورودی به مدل و ارتباط بین آن‌ها به‌منظور پیش‌بینی مقادیر جریان روزانه به‌وسیله مدل BN در شکل ۶ به‌عنوان نمونه برای الگوی ۴ ارائه شده است.

بر این اساس، در این مدل‌سازی جریان تا چهار تأخیر زمانی به‌عنوان داده‌های ورودی و جریان فردا به‌عنوان داده هدف است و هدف اصلی این مدل‌سازی یافتن بهترین مدل ورودی از دبی با تأخیرهای یک تا چهار روز برای پیش‌بینی دبی در یک روز آینده می‌باشد.

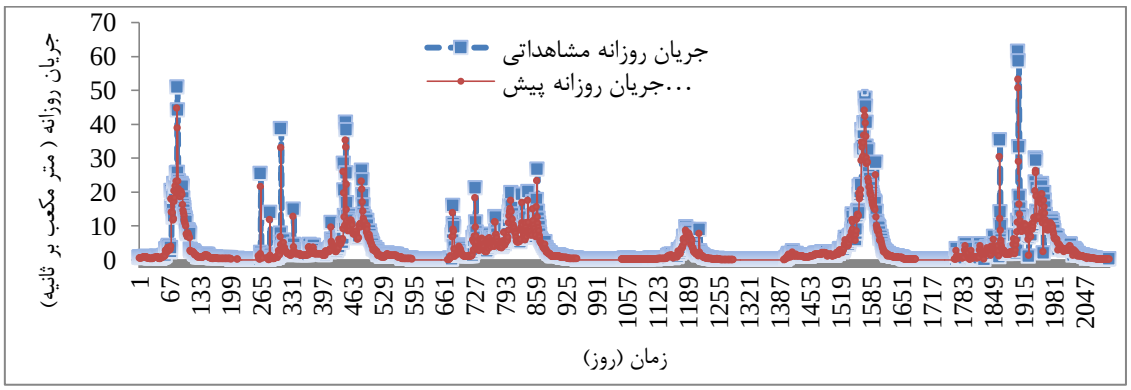


(الف) (ب)

شکل ۴- پراکنش داده‌های مشاهداتی و پیش‌بینی شده با روش GEP در الف) دوره آموزش و ب) دوره آزمون



(الف)



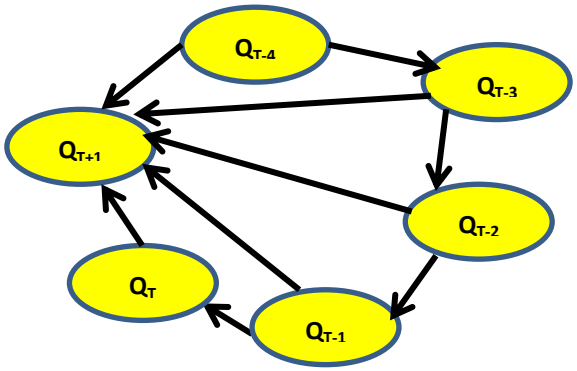
(ب)

شکل ۵- نمودار سری زمانی مقادیر مشاهداتی و پیش‌بینی شده با روش GEP، الف) مرحله آموزش و ب) مرحله آزمون

بر اساس معیارهای ارزیابی که در جدول ۳ ارائه شده است، مشاهده می‌شود که با افزایش تأخیرهای زمانی به‌عنوان الگوی ورودی، دقت پیش‌بینی مدل افزایش پیدا می‌کند، ولی با توجه به این‌که بهبود عملکرد مدل در گام زمانی سوم نسبت به گام زمانی چهارم ناچیز است، لذا الگوی برتر، الگوی ورودی ۳

می‌باشد. آماره‌های ارزیابی برای الگوی برتر BN در مرحله آموزش و آزمون به‌ترتیب دارای ضریب همبستگی ۰/۸۸۳ و ۰/۹۰۵ و RMSE شش و ۲/۶۷۹ و ضریب نش-ساتکلیف ۰/۷۷۷ و ۰/۸۱۷ می‌باشد که این مقادیر با نتایج مدل برتر روش GEP تقریباً برابر است که این نشان می‌دهد، هر دو مدل با دقت یکسانی مقادیر جریان روزانه را پیش‌بینی می‌کنند. اما با توجه به ساختار مدل‌سازی این دو روش مشهود است که روش شبکه بیزین به‌صورت ساده‌تر و پیچیدگی کمتر و با همان دقت روش GEP مقادیر جریان روزانه را پیش‌بینی می‌کند که نشان از برتری این مدل دارد. روش BN نیز همانند GEP رابطه ریاضی بین متغیرهای مستقل و وابسته را ارائه می‌دهد که از مزیت‌های مهم این روش محسوب می‌شود و از این طریق، می‌توان بدون استفاده مجدد از مدل و تنها با در دست داشتن مقادیر جریان با تأخیرهای مورد نظر مقادیر جریان برای فردا را پیش‌بینی کرد که این رابطه به‌صورت رابطه (۱۰) است. در شکل‌های ۷ و ۸ پراکنش و سری زمانی مقادیر پیش‌بینی شده و مشاهداتی در دو مرحله آموزش و آزمون با استفاده از روش BN ارائه شده است. بر اساس این نمودارها مشخص می‌شود مدل BN برای پیش‌بینی دبی روزانه از دقت قابل قبولی برخوردار است.

$$Q_{t+1} = 0.8038 \times Q_t , -0.0501 \times Q_{t-1} + 0.349 \times Q_{t-2} + 0.0863 \times Q_{t-3} + 0.0597 \times Q_{t-4}) \quad (۱۰)$$



شکل ۶- ساختار شبکه بیزین با چهار گام زمانی تأخیر برای پیش‌بینی جریان روزانه رودخانه مهاباد در زمان t+1

جدول ۳- معیارهای ارزیابی برای روش شبکه بیزین در دو مرحله آموزش و آزمون

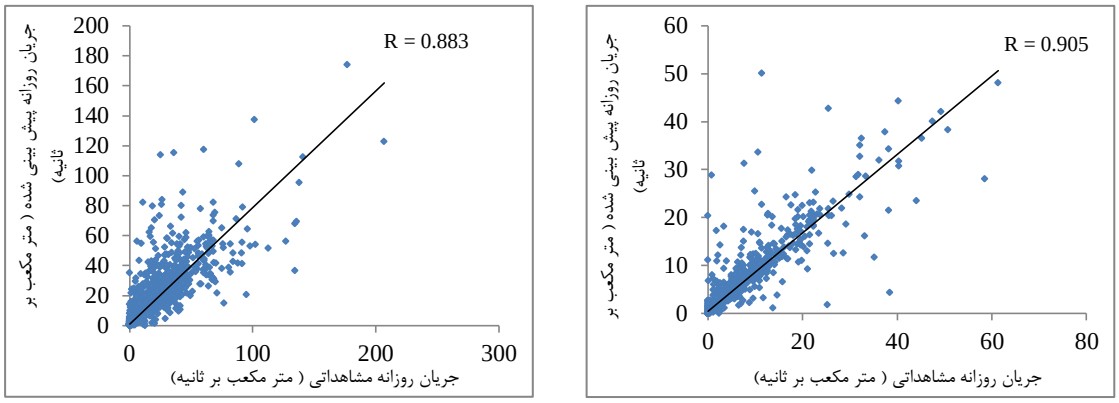
الگوی ورودی	دوره آموزش			دوره صحت سنجی		
	R	RMSE (m^3s^{-1})	E	R	RMSE (m^3s^{-1})	E
۱	۰/۷۵۴	۸/۱۵	۰/۶۴۳	۰/۸۰۹	۳/۸۷	۰/۷۲۲
۲	۰/۷۹۵	۷/۷۴	۰/۶۸۶	۰/۸۶۲	۳/۳۵۸	۰/۷۸۴
۳	۰/۸۸۳	۶	۰/۷۷۷	۰/۹۰۵	۲/۶۷۹	۰/۸۱۷
۴	۰/۸۹۱	۵/۹۴	۰/۷۸۳	۰/۹۰۸	۲/۶۳۵	۰/۸۲۴

نتیجه‌گیری

از آنجا که پیش‌بینی جریان رودخانه نقش مهمی در برنامه‌ریزی و مدیریت منابع آب دارد، هدف از این تحقیق، پیش‌بینی جریان رودخانه مهاباد به‌عنوان یکی از مهمترین رودخانه‌های حوزه آبخیز دریاچه ارومیه بوده است. بدین‌منظور، داده‌های جریان روزانه رودخانه مهاباد در یک دوره ۲۳ ساله و تا چهار تأخیر زمانی به‌عنوان ورودی مدل‌های پیش‌بینی شبکه بیزین و برنامه‌ریزی بیان‌ژن انتخاب شدند. بر این اساس، موارد زیر از این تحقیق قابل ارائه است:

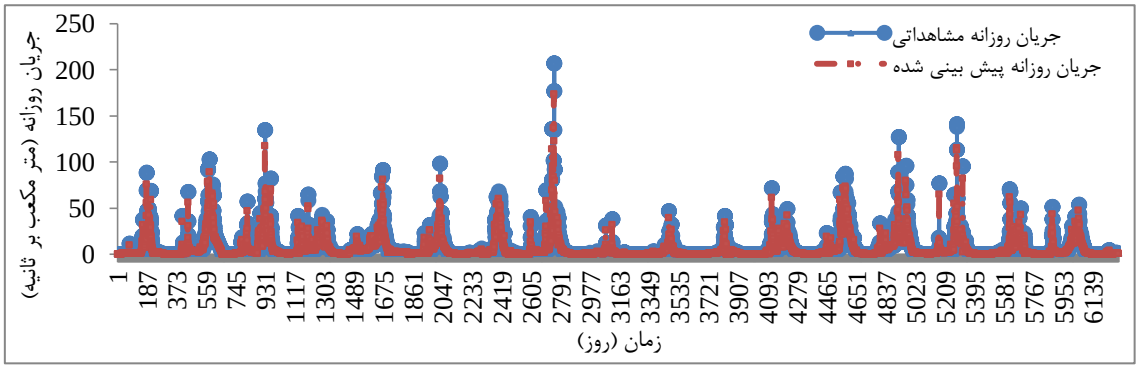
- مدل برتر در هر دو روش مدل با سه گام زمانی تأخیر است که در روش BN در دو بخش آموزش و آزمون دارای ضریب همبستگی ۰/۸۸۳ و ۰/۹۰۵ و RMSE به‌ترتیب شش و ۲/۶۷ متر مکعب بر ثانیه و در روش GEP نیز برای دو بخش آموزش و آزمون دارای ضریب همبستگی ۰/۸۸ و ۰/۹۰۲ و RMSE

۶/۰۶ و ۲/۷۱ متر مکعب بر ثانیه می‌باشد که نشان می‌دهد، هر دو مدل از دقت قابل قبول و تقریباً یکسان در پیش‌بینی مقادیر جریان روزانه مهاباد چای برخوردار هستند که این نتایج با تحقیقات Khu و همکاران (۲۰۰۱) و Guven (۲۰۰۹) مطابقت دارد.

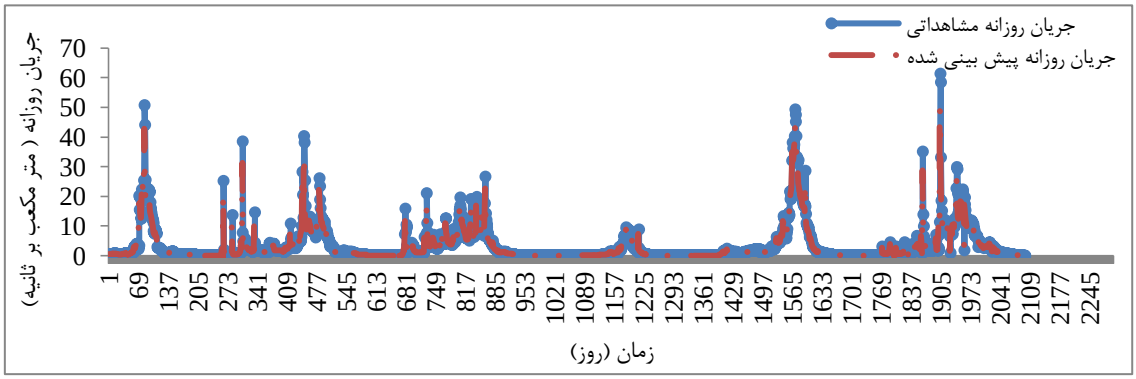


(الف) (ب)

شکل ۷- پراکنش داده‌های مشاهداتی و پیش‌بینی شده با شبکه بیزین، الف) دوره آموزش و ب) دوره آزمون



(الف)



(ب)

شکل ۸- نمودار سری زمانی مقادیر مشاهداتی و پیش‌بینی شده با شبکه بیزین در الف) مرحله آموزش و ب) مرحله آزمون

- از مزیت‌های هر دو روش BN و GEP ارائه رابطه ریاضی بین متغیرهای مستقل و وابسته است که برای یک روز آینده بدون نیاز به استفاده از مدل و نرم‌افزار می‌توان مقادیر جریان برای یک روز آینده را پیش‌بینی کرد که در روابط ارائه شده مدل شبکه بیزین از کلیه متغیرهای ورودی استفاده می‌کند، اما

روش GEP تنها سه متغیر جریان با تأخیر سه و چهار روز و بدون تأخیر را برای تشکیل رابطه ریاضی پیش‌بینی جریان پیشنهاد می‌کند.

- نتایج این تحقیق نشان داد، با توجه به این‌که هر دو روش از دقت نسبتاً یکسانی در پیش‌بینی مقادیر جریان روزانه برخوردار هستند، اما نحوه مدل‌سازی به روش BN به مراتب ساده‌تر از روش GEP می‌باشد و به‌عنوان یک روش کاربردی در مطالعات آتی می‌تواند مورد استفاده قرار گیرد.
- پیشنهاد می‌شود، این پژوهش با توجه شرایط مختلف جغرافیایی در مناطق دیگر تکرار و نتایج برای سایر رودخانه‌ها مورد آزمون قرار گیرد.

منابع مورد استفاده

1. Ahmadi, F., F. Radmanesh and R. Mirabbasi Najafabadi. 2016. Comparing the performance of support vector machines and bayesian networks in predicting daily river flow, case study: Barandoozchay River. Journal of Water and Soil Conservation, 22(6): 171-186 (in Persian).
2. Ahmadi, F., F. Radmanesh and R. Mirabbasi Najafabadi. 2016. Application of bayesian networks and genetic programming for predicting daily river flow, case study: Barandoozchay River. Irrigation Sciences and Engineering, 39(4): 213-223 (in Persian).
3. Awchi, T.A. 2014. River discharges forecasting in northern Iraq using different ANN techniques. Water Resources Management, 28(3): 801–814.
4. Box, G.E., G.M. Jenkins and G.C. Reinsel. 2011. Time series analysis: forecasting and control (Vol. 734). John Wiley and Sons, 712 pages.
5. Brandt, G. and H.J. Henriksen. 2003. Protection of drinking water sources for quality and quantity. Groundwater protection in the greater copenhagen area. In Future Scenarios for Water Management in Europe. FIRMA Conference, 19-20 February, Barcelona, SP.
6. Dorado J., J.R. Rabunal, A. Pazos, D. Rivero, A. Santos and J. Puertas. 2003. Prediction and modeling of the rainfall-runoff transformation of a typical urban basin using ANN and GP. Applied Artificial Intelligence, 17: 329-343.
7. Ferreira, C. 2001. Gene expression programming: a new adaptive algorithm for solving problems. Complex System, 13(2): 87–129.
8. Ghorbani, M.A. and R. Dehghani. 2016. Application of bayesian neural networks, support vector machines and gene expression programming analysis of monthly rainfall-runoff, case study: Kakareza River. Irrigation Sciences and Engineering, 39: 125-138 (in Persian).
9. Guven, A. 2009. Linear genetic programming for time-series modeling of daily flow rate. Journal of Earth System Science, 118(2): 157-173.
10. He, Z., X. Wen, H. Liu and J. Du. 2014. A comparative study of artificial neural network, adaptive neuro fuzzy inference system and support vector machine for forecasting river flow in the semi-arid mountain region. Journal of Hydrology, 509: 379-386.
11. Khu, S.T., S.Y. Liong, V. Babovic, H. Madsen and N. Muttil. 2001. Genetic programming and its application in real-time runoff forming. Journal of American Water Resources Association, 37(2): 439-451.
12. Neapolitan, R.E. 2003. Learning Bayesian networks. Prentice Hall Series in Artificial Intelligence, 693 pages.
13. Reggiani, P. and A.H. Weerts. 2008. A Bayesian approach to decision-marking under uncertainty: an application to real-time forecasting in the river Rhine. Journal of Hydrology, 356: 56-69.
14. Rezaei, E., B. Shahinejad and H. Yonesi. 2019. Analysis and evaluation of effective parameters on the amount of total dissolved solids in rivers. Watershed Engineering and Management, 11(1): 147-165 (in Persian).
15. Terzi, O. and G. Ergin. 2014. Forecasting of monthly river flow with autoregressive modeling and data-driven techniques. Neural Computing and Applications, 25(1): 179-188.
16. Zhang, H., V.P. Singh, B. Wang and Y. Yu. 2016. CEREF: a hybrid data-driven model for forecasting annual streamflow from a socio-hydrological system. Journal of Hydrology, 540: 246-256.

DOI: 10.22092/ijwmse.2021.126342.1656

Daily river flow estimation based on intelligent models, case study: Mahabad River

Abbas Abbasi[*1], Keivan Khalili[2], Javad Behmanesh[3] and Akbar Shirzad[4]
[1*] PhD, Faculty of Agriculture, Urmia University, Iran, [2] Assistant Professor, Faculty of Agriculture, Urmia University, Iran, [3] Professor, Faculty of Agriculture, Urmia University, Iran and [4] Associate Professor, Faculty of Environment, Urmia University of Technology, Iran

Received: 18 May 2020 Accepted: 15 March 2021

Abstract

The correct and accurate estimation of river flow can play an important role in reducing the effects of flood damage. In this research, Gene Expression Programming (GEP) model and Bayesian Network (BN) were used to predict daily flow of Mahabad River in Urmia Lake Basin. Accordingly, four input models with a delay of one to four days used to estimate daily flow at time t+1 over a 23-years period and 75% of data was used to train the models and 25% of the remaining data was used for the test stage. Results showed that the best model in both methods was the input pattern with three-time lags. Also, based on the correlation coefficient (R), Root Mean Square Error (RMSE) and Nash-Sutcliffe (E) coefficient in the test stage of the GEP method with R=0.902, RMSE=2.71(m^3s^{-1}) and E=0.812 compared to the BN method with R=0.905, RMSE=2.679(m^3s^{-1}) and E=0.817 is more accurate. In general, both methods have acceptable accuracy and are they relatively similar, but because of the simpler modeling, Bayesian Network method can be used as an efficient method for predicting river flow.

Keywords: Bayesian Network, Flood, Gene Expression Programming, Prediction, Urmia Lake

[*] Corresponding author: abasi.abas2010@gmail.com