نشریه آب و خاک (علوم و صنایع کشاورزی)
جلد ۲۹، شماره ۵، آذر – دی ۱۳۹۴، ص. ۱۲۰۶-۱۱۹۰

Journal of Water and Soil
Vol. 29, No. 5, Nov.-Dec. 2015, p. 1190-1206

پیش‌بینی جریان رودخانه شهرچای در حوضه آبریز دریاچه ارومیه با استفاده از برنامه‌ریزی ژنتیک و مدل درختی M5

سعید صمدیان فرد[1*] – رضا دلیرحسن نیا[2]

تاریخ دریافت: ۱۳۹۳/۰۱/۲۳
تاریخ پذیرش: ۱۳۹۴/۰۱/۳۰

چکیده

منابع آب شیرین قابل استحصال با محدودیت جدی مواجه است، بنابراین پیش‌بینی هرچه دقیق‌تر جریان رودخانه در تحلیل بسیاری از پدیده‌های خشکسالی و سیلاب، آبگیری از رودخانه‌ها و سایر مسائل مرتبط از اهمیت بالایی برخوردار بوده و از ارکان اساسی برنامه‌ریزی و مدیریت منابع آب‌های سطحی می‌باشد. از این‌رو متخصصان همواره برای تخمین صحیح دبی رودخانه و اصلاح روش‌های موجود تلاش می‌نمایند. در این راستا و در تحقیق حاضر، از روش‌های هوشمند برنامه‌ریزی ژنتیک و مدل درختی M5 برای مدل‌سازی و پیش‌بینی جریان رودخانه شهرچای در حوضه آبریز دریاچه ارومیه استفاده شده است. بدین منظور، داده‌های میانگین ماهانه دبی رودخانه شهرچای در ایستگاه بند در بازه زمانی بین سال‌های ۱۳۳۰ الی ۱۳۹۰ برای واسنجی و صحت‌سنجی روش‌های مذکور مورد استفاده قرار گرفته و دقت این روش‌ها با استفاده از پارامترهای آماری جذر میانگین مربعات خطا، میانگین خطای مطلق و ضریب همبستگی مورد بررسی قرار گرفته است. نتایج حاصل از این مطالعه نشان دادند که روش برنامه‌ریزی ژنتیک با دارا بودن خطای ۳/۳۰۹۴ و مدل درختی M5 با خطای ۳/۵۵۱۴ در حالت استفاده از حافظه‌های دبی یک، دو و سه ماه قبل (Q_{t-1}, Q_{t-2}, Q_{t-3}) با داشتن کمترین مقدار خطا، عملکرد مناسبی در پیش‌بینی جریان رودخانه داشته‌اند. در نهایت روش برنامه‌ریزی ژنتیک در حالت استفاده از توابع ریاضی متشکل از چهار عملی اصلی، لگاریتم و توان و با در نظر گرفتن پارامترهای ورودی Q_{t-1},Q_{t-2},Q_{t-3} و دارا بودن بهترین عملکرد، به‌عنوان روشی مناسب برای پیش‌بینی جریان رودخانه پیشنهاد گردید.

واژه‌های کلیدی: پارامترهای آماری، تخمین، دبی جریان، روش‌های هوشمند

مقدمه

پیش‌بینی دقیق دبی جریان نکته کلیدی در برنامه‌ریزی و مدیریت بهینه منابع آب به شمار می‌آید. افزایش برای تقاضای آب در مناطق مختلف به‌ویژه در ناحیه‌های خشک و نیمه‌خشک، نیاز به مدیریت بهینه منابع آب را بیش از پیش نشان می‌دهد. بر این پایه دست‌یابی به روش‌های مطمئن پیش‌بینی جریان رودخانه‌ها به‌منظور برنامه‌ریزی در بهره‌برداری به‌موقع از منابع آب اهمیت روزافزونی دارد. در سال‌های اخیر، کاربرد روش‌های هوشمند از جمله شبکه‌های عصبی مصنوعی، منطق فازی و برنامه‌ریزی ژنتیک در زمینه‌هایی که در آن‌ها ارتباط بین پارامترهای ورودی و خروجی غیرخطی بوده، به‌صورت فزاینده‌ای گسترش پیدا کرده‌اند. این روش‌ها به‌عنوان یک جعبه سیاه مناسب که کمتر در قید و بند مسائل فیزیکی بوده و قادرند فرآیندهای غیرخطی و غیرایستای جریان رودخانه را بدون نیاز به عامل‌های محیطی و هندسی مؤثر بر جریان رودخانه مدل‌سازی کنند، می‌باشند (۸).

در زمینه کاربرد روش‌های هوشمند در مطالعات منابع آب پژوهش‌های متعددی گزارش شده است، به‌عنوان نمونه نورانی و صالحی (۲۳) با استفاده از سه روش شبکه عصبی، استنتاج فازی و شبکه تطبیقی عصبی- فازی مدل بارش-رواناب حوضه لیقوان‌چای واقع در استان آذربایجان‌شرقی را ارائه نمودند. آن‌ها به این نتیجه رسیدند که شبکه عصبی و منطق فازی، به‌تنهایی قادر به تبیین این مدل نمی‌باشد، ولی در صورت تلفیق این دو روش، نتیجه مطلوب برای مدل‌سازی بارش – رواناب در سطح حوضه‌ی مورد بررسی حاصل می‌شود. فربودنام و همکاران (۱۰) از روش‌های برنامه‌ریزی ژنتیک و شبکه‌های عصبی مصنوعی جهت پیش‌بینی جریان روزانه رودخانه لیقوان در حوضه آبریز دریاچه ارومیه در دوره آماری ۱۳۷۶ تا ۱۳۸۰ استفاده نموده و اظهار داشتند که روش برنامه‌ریزی ژنتیک از دقت بیشتری نسبت به مدل شبکه‌های عصبی مصنوعی برخوردار است.

۱ و۲- استادیاران گروه مهندسی آب، دانشکده کشاورزی، دانشگاه تبریز
(*- نویسنده مسئول: Email: s.samadian@tabrizu.ac.ir)

ابراهیمی‌محمدی و بشری سه‌قلعه (۷) در مدل‌سازی تغییرات دبی ماهانه رودخانه قره‌سو با استفاده از سری‌های زمانی و شبکه‌های عصبی مصنوعی و با به کارگیری داده‌های بارندگی، دما، تبخیر و دبی به این نتیجه رسیدند که نتایج حاصل از پیش‌بینی بر اساس مدل شبکه عصبی مصنوعی انطباق بیشتری با مقادیر مشاهداتی دارد. نبی‌زاده و همکاران (۲۱) در پژوهشی از مدل‌های مبتنی بر منطق فازی شامل استنتاج فازی و استنتاج فازی – عصبی تطبیقی به‌منظور پیش‌بینی جریان رودخانه لیقوان چای استفاده نمودند. ارزیابی نتایج پیش‌بینی‌ها با استفاده از معیارهای آماری نشان داد که مدل استنتاج فازی – عصبی تطبیقی با دقت بالاتر و پراکندگی کمتری نسبت به مدل استنتاج فازی دبی این رودخانه را پیش‌بینی کرده است. ناوه و همکاران (۲۲) جریان رودخانه‌های باراندوزچای و شهرچای ارومیه را با استفاده از مدل غیرخطی سری زمانی دو–خطی پیش‌بینی نموده و نتایج آن را با نتایج حاصل از روش آرما (ARMA) مقایسه نمودند. نتایج این تحقیق نشان داد که در دو رودخانه مورد مطالعه، مدل غیرخطی استفاده شده به دلیل خطای کمتر نسبت به مدل‌های خطی آرما، کارایی بهتری داشته است. مرادی‌زاده کرمانی و همکاران (۲۰) قابلیت‌های نظریه آشوب و برنامه‌ریزی ژنتیک را در تخمین دبی روزانه رودخانه لیقوان مورد مطالعه قرار دادند. نتایج به دست آمده حاکی از دقت هر دو روش فوق‌الذکر در مدل‌سازی و پیش‌بینی دبی رودخانه می‌باشد. قبادیان و همکاران (۱۲) عملکرد روش برنامه‌ریزی بیان ژن در روندیابی سیلاب رودخانه زنگمار را با روش موج دینامیکی مقایسه نمودند. نتایج روندیابی شش هیدروگراف سیل، بین سال‌های ۱۳۸۴ تا ۱۳۸۸ نشان داد که مدل برنامه‌ریزی بیان ژن قادر است با دقت بیشتری حجم هیدروگراف خروجی را پیش‌بینی کند. همچنین نتیجه گرفتند که در خصوص دبی پیک و زمان وقوع آن، مدل موج دینامیکی برتری دارد. خزایی و میرزایی (۱۵) پژوهشی را با هدف مقایسه مدل‌های مختلف شبکه عصبی مصنوعی (MLP و RBF) و سری‌های زمانی آرما در برآورد دبی ماهانه در حوزه آبخیز طالقان برای یک دوره ۳۰ ساله بین سال‌های ۱۳۵۶ تا ۱۳۸۶ به انجام رساندند. آن‌ها در روش شبکه عصبی مصنوعی از توابع محرک سیگموئیدی و پارامترهای تعداد تکرار، ضریب یادگیری، تعداد نرون مخفی و خطای هدف و همچنین در روش آرما از بین مدل‌های مختلف روشی که دارای کمترین میزان خطا بود را به‌عنوان مدل بهینه انتخاب نمودند. در مجموع دقت‌سنجی مدل‌ها براساس آماره‌های ریشه میانگین مربعات خطا و ضریب همبستگی حاکی از دقت بیشتر شبکه عصبی مصنوعی نسبت به مدل‌های سری زمانی می‌باشد. خلیلی و همکاران (۱۳) الگوی تلفیقی BL-ARCH را به‌منظور پیش‌بینی دبی رودخانه شهرچای ارومیه پیشنهاد نمودند. در این پژوهش از داده‌های روزانه ۳۱ ساله دبی رودخانه شهرچای ارومیه واقع در غرب دریاچه ارومیه و استان آذربایجان غربی، برای پیش‌بینی ۱۱ ساله جریان رودخانه استفاده شده است. نتایج حاصله نشان داد که مدل تلفیقی پیشنهادی با خطای ۴/۵۲، نسبت به مدل غیرخطی سری زمانی دو– خطی با خطای ۶/۷۷ برازش بهتری دارد. آن‌ها بیان نمودند که مدل پیشنهادی می‌تواند در پیش‌بینی جریان‌های کوتاه مدت رودخانه به کارگرفته شود.

خو و همکاران (۱۶) در تحقیقی در مورد حوضه آبریز اورگوال[1] در کشور فرانسه، از برنامه‌ریزی ژنتیک برای پیش‌بینی رواناب بهره برده و نتایج حاصل را با مقادیر مشاهداتی و نیز مقادیر محاسبه‌شده توسط روش‌های کلاسیک مورد مقایسه قرار دادند. نتایج حاصل از این تحقیق، بیانگر دقت قابل قبول برنامه‌ریزی ژنتیک می‌باشد. عقیل و همکاران (۳) برای مدل‌سازی رفتار ساعتی و روزانه رواناب، مطالعه مقایسه‌ای بین مدل‌های شبکه عصبی مصنوعی و مدل‌های عصبی– فازی انجام دادند. آن‌ها به این نتیجه رسیدند که مدل عصبی– فازی عملکرد بهتری در مدل‌سازی رواناب نسبت به مدل‌های شبکه عصبی مصنوعی دارد. فیرات و گانگر (۱۱) مدل عصبی– فازی را برای پیش‌بینی جریان رودخانه بزرگ مندرس در ترکیه مورد ارزیابی قرار داده و به این نتیجه رسیدند که مدل عصبی – فازی می‌تواند با موفقیت، صحت و قابلیت اطمینان بالا برای پیش‌بینی جریان رودخانه به کار رود. آیتک و آلپ (۴) برای مدل‌سازی بارش – رواناب در حوضه رودخانه جونیاتا[2] در پنسیلوانیای آمریکا از دو روش شبکه عصبی مصنوعی و برنامه‌ریزی ژنتیک استفاده کردند. مقایسه نتایج نشان داد که روش برنامه‌ریزی ژنتیک در مقایسه با شبکه عصبی مصنوعی عملکرد بهتری داشته است. خطیبی و همکاران (۱۴) عملکرد سه روش برنامه‌ریزی ژنتیک، شبکه عصبی و نروفازی را در روندیابی سیلاب رودخانه قزل ایرماق ترکیه مورد ارزیابی قرار دادند. نتایج مطالعه آن‌ها نشان داد که مدل برنامه‌ریزی ژنتیک با دقت بیشتری هیدروگراف خروجی را شبیه‌سازی می‌کند. لوندهه و دیکسیت (۱۹) مدل درختی M5 را برای پیش‌بینی جریان روزانه رودخانه در دو ایستگاه در حوضه‌های آبریز نارمادا و کریشنا در هندوستان به کار بردند. آن‌ها گزارش نمودند که نتایج حاصل از مدل درختی تطابق معقولی با داده‌های مشاهداتی داشته است. ستاری و همکاران (۲۶) توانایی مدل‌های درختی M5 و بردارهای ماشین در پیش‌بینی جریان رودخانه سوهو در ترکیه را مورد مطالعه قرار داده و نشان دادند که پیش‌بینی‌های مدل درختی M5 تطابق بهتری با داده‌های مشاهداتی داشته است.

هدف پژوهش حاضر، مقایسه عملکرد مدل‌های هوشمند برنامه‌ریزی ژنتیک و مدل درختی M5 در پیش‌بینی جریان رودخانه شهرچای در حوضه آبریز دریاچه ارومیه و به دست آوردن بینش کلی

1- Orgeval catchment
2- Juniata river basin

نسبت به توانایی هر کدام از روش‌های مذکور و همچنین ارائه روابط ریاضی صریح برای محاسبه و پیش‌بینی آسان‌تر جریان رودخانه می‌باشد.

منطقه مورد مطالعه

رودخانه شهرچای به‌عنوان منبع اصلی تأمین کننده آب شرب شهر ارومیه و حومه آن، آب کشاورزی مورد نیاز اراضی تحت پوشش و همچنین به‌عنوان یکی از ورودی‌های مهم دریاچه ارومیه دارای اهمیت زیادی می‌باشد. این رودخانه از کوه‌های زرینه با ارتفاع ۳۱۰۰ متر، کمال با ارتفاع ۳۳۸۶ متر و زریناتابوتان با ارتفاع ۳۳۶۹ متر، واقع در ۴۴ کیلومتری جنوب‌غربی شهرستان ارومیه سرچشمه گرفته و با دریافت شاخه‌هایی که همگی از کوه‌های مرزی ایران و ترکیه سرچشمه می‌گیرند، وارد منطقه برده سور می‌شود. سپس با عبور از روستای بند وارد شهر ارومیه می‌شود. از میان شهر گذشته و بعد از مشروب ساختن روستاهای متعدد سر راه خود بعد از روستای کشتیبان در محلی به نام دماغه حصار وارد دریاچه ارومیه می‌شود. جهت غالب رودخانه غرب به شرق می‌باشد (شکل ۱).

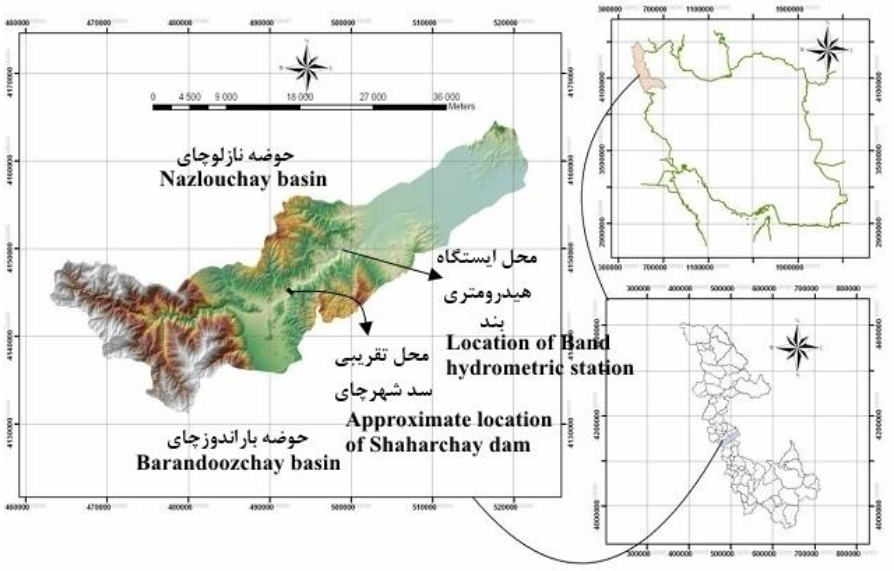
شاید در نگاه اول چنین به نظر رسد که وجود سد مخزنی شهرچای در فاصله تقریبی ۱۲ کیلومتری در بالادست ایستگاه هیدرومتری بند، جریان رودخانه را تحت تأثیر قرار می‌دهد، اما به نظر نگارندگان فاصله تقریبا زیاد محل اندازه‌گیری جریان و سد مذکور و همچنین اضافه شدن جریان‌های محلی به آب رودخانه، رژیم رودخانه را مستقل از تأثیر سد می‌نماید.

بدیهی است که همواره بر روی رودخانه‌هایی با جریان قابل توجه، انواع تأسیسات آبگیری از جمله بند انحرافی، سد و ایستگاه پمپاژ احداث می‌شود. اما شاهد هستیم که تحقیقات بسیاری در زمینه پیش‌بینی جریان با انواع مدل‌ها با موفقیت به انجام می‌رسد. بر این اساس هدف از انجام پژوهش حاضر بررسی، مدل‌سازی و پیش‌بینی روند موجود رودخانه می‌باشد.

مواد و روش‌ها

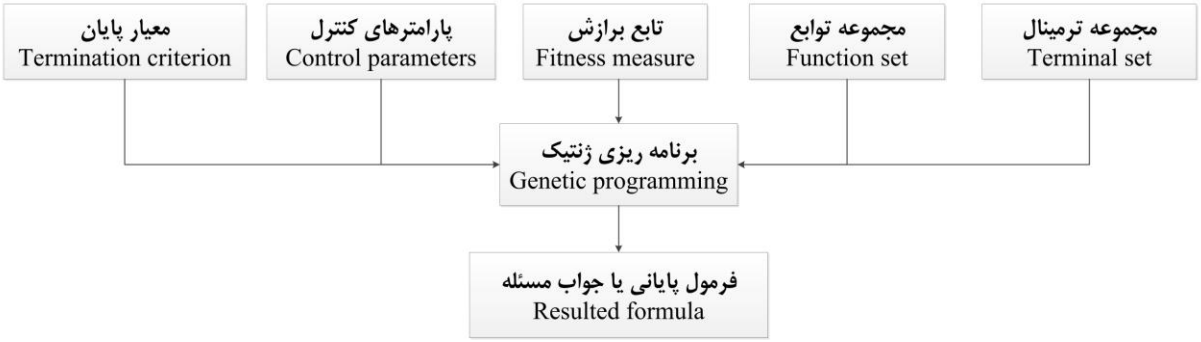
برنامه‌ریزی ژنتیک

برنامه‌ریزی ژنتیک تعمیم یافته الگوریتم ژنتیک می‌باشد که برای اولین بار توسط کزا (۱۷) بر اساس تئوری داروین ارائه شد. به این ترتیب که جمعیتی در جهت تکامل به‌صورت انتخابی، جمعیت نامناسب را رها کرده و فرزندانی اصلاح شده ایجاد می‌کنند. برنامه‌ریزی ژنتیک یک تکنیک برنامه‌ریزی خودکار می‌باشد که راه حل مسئله را با استفاده از برنامه کامپیوتری ارائه می‌کند. در این روش در ابتدای فرآیند هیچ‌گونه رابطه تابعی در نظر گرفته نشده و این روش قادر به بهینه‌سازی ساختار مدل و مولفه‌های آن می‌باشد. برنامه‌ریزی ژنتیک بر خلاف الگوریتم ژنتیک روی ساختار درختی فرمول‌ها به‌جای سلسله ارقام دودوئی عمل می‌کند. ساختارهای درختی از مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول‌ها) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند.



شکل ۱- موقعیت حوضه‌های استان آذربایجان غربی و حوضه آبریز شهرچای و ایستگاه هیدرومتری بند

Figure 1- Location of the watersheds in west Azerbaijan and Shaharchay watershed and Band hydrometric station



شکل ۲- طرح کلی گام‌های مقدماتی برنامه‌ریزی ژنتیک

Figure 2- Schematic of preliminary steps in genetic programming

تحقیقات گذشته نشان می‌دهد که در استفاده از محاسبات نرم (مثل برنامه‌ریزی ژنتیک و شبکه‌های عصبی و ...) به بعضی از مشکلات مدل‌های فیزیکی غلبه شده و این روش‌ها می‌توانند یک راه مؤثر و کارآمد برای مدل کردن فرآیندهای فیزیکی در حالت فقدان داده کافی در دسترس، باشند. قبل از مراحل اجرایی برنامه‌ریزی ژنتیک گام‌های مقدماتی زیر باید توسط کاربر تعیین شوند (۱۷).

مجموعه ترمینال‌ها (متغیرهای مسئله، اعداد ثابت تصادفی)

مجموعه عملگرهای ریاضی مورد استفاده در فرمول‌ها

انتخاب تابع برازش مناسب (RRSE, MSE, R,...)، برای سنجش برازش فرمول‌ها

تعیین پارامترهای کنترل‌کننده اجرای برنامه (اندازه جمعیت، احتمال مربوط به‌کارگیری عمل‌های ژنتیکی و جزئیات دیگر مربوط به اجرای برنامه)

معیار پایان و ارائه نتایج اجرای برنامه (مثل، تعداد تولید جمعیت جدید، تعیین یک مقدار مشخص برای برازش فرمول‌ها که اگر میزان برازش برابر یا بیشتر از آن مقدار شد، اجرا متوقف شود).

در شکل ۲ طرح کلی گام‌های مقدماتی برنامه‌ریزی ژنتیک نشان داده شده است.

فرآیند اجرایی گام به گام برنامه‌ریزی ژنتیک به صورت زیر است:

تولید یک جمعیت اولیه از فرمول‌ها که این فرمول‌ها از ترکیب تصادفی مجموعه توابع (عملگرهای ریاضی مورد استفاده در فرمول‌ها) و ترمینال‌ها (متغیرهای مسئله و اعداد ثابت) ایجاد می‌شوند.

هر یک از افراد جمعیت مذکور با استفاده از توابع برازش مورد ارزیابی قرار می‌گیرند.

تولید یک جمعیت جدید از فرمول‌ها، که مراحل زیر برای تولید یک جمعیت جدید دنبال می‌شود:

الف) یکی از عملگرهای ژنتیکی تلاقی، جهش و تولید مثل انتخاب می‌شود (این سه عملگر ژنتیکی، مهم‌ترین عملگرهای ژنتیکی مورد استفاده در برنامه‌ریزی ژنتیک می‌باشند. عملگرهای دیگری مثل اصلاح ساختار و... نیز با احتمال کمتر مورد استفاده قرار می‌گیرند). ب) تعداد مناسبی از افراد جمعیت حاضر انتخاب می‌شوند (انتخاب فرد یا افرادی از جمعیت مذکور به‌صورت احتمالاتی می‌باشد که در این انتخاب احتمالاتی منفردهای با برازش بهتر به منفردهای نامرغوب ترجیح داده می‌شوند و این بدان معنی نیست که حتما منفردهای نامرغوب حذف می‌شوند). ج) از عملگر ژنتیکی انتخاب شده برای تولید فرزند (فرمول جدید) استفاده می‌شود.

د) فرزند (فرمول جدید) تولید شده در یک جمعیت جدید وارد می‌شود.

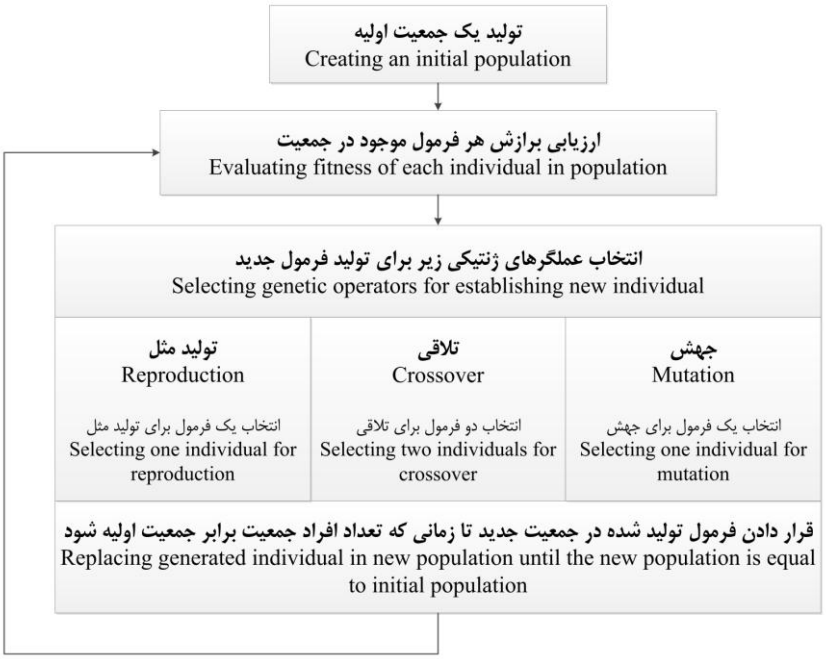
ه) مدل مورد نظر با استفاده از تابع برازش مورد ارزیابی واقع می‌شود.

گام سوم تا نیل به حداکثر تعداد تولید، تکرار خواهد شد. شکل کلی گام‌های اجرایی برنامه‌ریزی ژنتیک در شکل ۳ و پارامترهای مورداستفاده در جدول ۱ ارائه‌شده‌اند.

مدل درختی M5

مدل درختی M5 (۲۵) زیر مجموعه‌ای از روش‌های یادگیری ماشینی و داده‌کاوی است. داده‌کاوی به فرآیند جست و جو و کشف مدل‌های گوناگون، مختصرسازی‌ها و اخذ مقادیر از مجموعه‌ای از مقادیر معلوم اطلاق می‌گردد (۱). روش‌های داده‌کاوی برای مجموعه داده‌های بزرگ با متغیرهای زیاد ساخته شده‌اند، بنابراین از روش‌های آماری قدیمی که برای مجموعه داده‌های کوچک با متغیرهای اندک طراحی شده‌اند تفاوت دارند.

روش‌های مبتنی بر درخت تصمیم به‌عنوان یکی از شناخته‌شده ترین تکنیک‌های داده‌کاوی، ویژگی هدف را به‌عنوان خروجی به‌صورت یک مدل با سازه درختی با استفاده از داده‌های ورودی پیش‌بینی و یا طبقه‌بندی می‌نمایند. مدل M5 یک مدل درختی برای پیش‌بینی صفات عددی پیوسته است که در آن توابع رگرسیونی خطی در برگ‌های این درخت متظاهر می‌شوند (۲۴) که در سال‌های اخیر تحول قابل توجهی در مسائل طبقه‌بندی و پیش‌بینی ایجاد نموده است.



شکل ۳- مروری بر شکل کلی گام‌های اجرایی برنامه‌ریزی ژنتیک

Figure 3- Schematic of operational steps in genetic programming

جدول ۱- پارامترهای مورد استفاده در برنامه‌ریزی ژنتیک

Table 1- Parameters which is used in genetic programming

پارامتر Parameter	مقدار Value
توابع مورد استفاده Used functions	+, -, ×, /, √, Ln(x), exp, Power, Sin, Cosine, Arctangent
تعداد کروموزوم‌ها Number of chromosomes	30
تعداد ژن‌ها Number of genes	3
عملگر ریاضی بین ژن‌ها Mathematical operator between genes	جمع (+) Addition
سرعت جهش Mutation	0.044
سرعت وارونگی Inversion	0.1
سرعت تلاقی با یک نقطه One-point recombination	0.3
سرعت تلاقی با دو نقطه Two-point recombination	0.3
سرعت تلاقی ژن Gene recombination	0.1
سرعت جابه جایی Gene transposition	0.1

نتایج مدل درختی برای فهم و شبیه‌سازی آسان هستند و خروجی مدل از دقت بالایی برخوردار است که می‌توان آن را با سایر مدل‌ها مقایسه کرد. مدل درختی می‌تواند در مسائل گوناگون مورداستفاده محققین قرار گیرد (۹).

یک درخت تصمیم معمولا از چهار بخش ریشه، شاخه، گره‌ها و برگ‌ها تشکیل شده است که گره‌ها با دایره نشان داده می‌شوند و شاخه‌ها نشان‌دهنده اتصال بین گره‌ها می‌باشند. درخت تصمیم به‌منظور سادگی در رسم معمولا از چپ به راست و یا از بالا به پایین کشیده می‌شود، به‌طوری‌که ریشه (گره اول) در بالا قرار می‌گیرد. انتهای یک زنجیره را برگ می‌نامند. هر گره مربوط به یک خصوصیت معین است و شاخه‌ها به معنای بازه‌ای از مقادیر هستند، این بازه‌های مقادیر بخش‌های مختلف مجموعه مقادیر معلوم را برای خصوصیت‌ها به دست دهند. عمل انشعاب توسط یکی از متغیرهای پیش‌بینی کننده انجام می‌پذیرد، بازه‌های انشعاب طوری انتخاب می‌شوند که مجموع مجذور انحراف از میانگین داده‌های هر گره را حداقل کنند (۹).

اولین مرحله برای ایجاد یک مدل درختی، استفاده از یک معیار انشعاب است. معیار انشعاب برای الگوریتم M5 بر اساس عملکرد انحراف استاندارد مقادیر هر کلاس و یا طبقه است که در هر گره به‌دست‌آمده است. این روش اساس روش‌های طبقه‌بندی است که آنتروپی نامیده می‌شود. آنتروپی می‌تواند به‌عنوان معیار میزان آشفتگی و بی‌نظمی یک سیستم تفسیر شود. معیار انشعاب بیانگر میزان خطا در آن گره می‌باشد و مدل حداقل خطای مورد انتظار را به‌عنوان نتیجه‌ی آزمایش هر صفت در آن گره محاسبه می‌کند. خطای مدل عموما با اندازه‌گیری دقت پیش‌بینی مقادیر هدف موارد دیده نشده سنجش می‌شود. فرمول محاسبه کاهش انحراف استاندارد[1] (SDR) به‌صورت زیر می‌باشد.

$$SDR = sd(T) - \sum_{i=1}^{n} \frac{|T_i|}{|T|} sd(T_i) \qquad (۱)$$

که در آن، T مجموعه‌ای از نمونه‌هاست که به هر گره وارد می‌شود، Ti زیر مجموعه‌ای از نمونه‌هاست که i مین نتیجه آزمون بالقوه را دارند، sd انحراف معیار و n تعداد نمونه‌ها می‌باشند (۱).

مدل‌های تصمیم درختی انواع مختلف دارد. هنگامی‌که خروجی یک درخت، یک مجموعه گسسته از یک مجموعه مقادیر ممکن است، به آن طبقه‌بندی درختی گفته می‌شود. هنگامی‌که بتوان خروجی درخت را یک عدد حقیقی در نظر گرفت، آن را رگرسیون درختی می‌نامند. به‌عبارت‌دیگر، اگر متغیرهای ورودی به سیستم عددی باشند، از رگرسیون درختی و اگر مجموعه‌ای از مقادیر[2] باشند، از طبقه‌بندی درختی استفاده می‌شود (۹). فرآیند انشعاب در هر گره بارها تکرار می‌گردد تا به گره پایانی (برگ) برسد که در برگ، مجموع مربعات انحراف از میانگین داده‌ها حدوداً به صفر می‌رسد. با این کار درخت بزرگی توسعه پیدا خواهد کرد. کار با این درخت بزرگ که شاخه‌ها و گره‌های زیادی دارد سخت می‌باشد، بنابراین برای رسیدن به یک درخت بهینه و کارآمد باید شاخه‌های اضافی درخت هرس شود. دو روش برای هرس کردن درخت وجود دارد: ۱) هرس قبل از شکل‌گیری درخت حداکثر (درخت توسعه یافته) ۲) هرس بعد از شکل‌گیری درخت حداکثر. در روش اول فرآیند هرس اجازه نمی‌دهد شاخه‌های اضافی تولید شوند، ولی در روش دوم ابتدا درخت حداکثر تشکیل می‌شود، سپس فرآیند هرس انجام می‌گیرد. درخت بهینه بر اساس حداقل کردن خطای پیش‌بینی انتخاب می‌شود (۹). پس از هرس کردن، فرآیند صاف کردن برای جبران گسیختگی‌های تند بین مدل‌های خطی هم‌جوار در برگ‌های درخت هرس شده انجام می‌شود (۵).

معیارهای ارزیابی دقت روش‌های مورد استفاده

مقادیر خطای بین سه روش محاسباتی و داده‌های مشاهداتی توسط جذر میانگین مربعات خطا (RMSE)، میانگین خطای مطلق (MAE) و ضریب همبستگی (R) و با استفاده از روابط زیر مورد بررسی قرار گرفت (۱۸).

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - y_i)^2} \qquad (۲)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}|x_i - y_i| \qquad (۳)$$

$$R = \frac{\left(\sum_{i=1}^{n} x_i y_i - \frac{1}{n}\sum_{i=1}^{n} x_i \sum_{i=1}^{n} y_i\right)}{\left(\sum_{i=1}^{n} {x_i}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)^2\right)\left(\sum_{i=1}^{n} {y_i}^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2\right)} \qquad (۴)$$

در روابط ۲ تا ۴، x_i و y_i مقادیر دبی جریان مشاهداتی و پیش‌بینی شده و n تعداد مشاهدات می‌باشند.

روش تحقیق

به‌منظور رسیدن به اهداف تحقیق، ابتدا داده‌های میانگین ماهانه دبی رودخانه شهرچای در ایستگاه بند در طول سال‌های ۱۳۹۰-۱۳۳۰ تهیه گردید. سپس دو سوم داده‌ها برای واسنجی و یک سوم باقیمانده برای صحت‌سنجی روش‌های مورد مطالعه شامل برنامه‌ریزی ژنتیک و مدل درختی M5 مورد استفاده قرار گرفت. لازم

1- Standard deviation reduction
2- Finite set of values

به ذکر است که تقسیم داده‌ها به دوسوم و یک‌سوم به‌صورت تصادفی بوده و قبل از مرتب‌سازی داده‌ها انجام گرفته است. علاوه بر این، تقسیم‌بندی دوسوم – یک‌سوم حالت متداولی است که توسط اکثریت محققان مورد استفاده قرار می‌گیرد. چنانچه پیش‌تر اشاره گردید، انتخاب یک‌سوم یا یک‌دهم به‌صورت تصادفی بوده و این نوع انتخاب نتایج را تحت تأثیر قرار نمی‌دهد. همچنین داده‌های ماهانه از روند منظم و تکرارپذیر برخوردارند و قابلیت مدل‌سازی بالایی دارند اما در داده‌های روزانه این نظم و روند مشاهده نمی‌گردد.

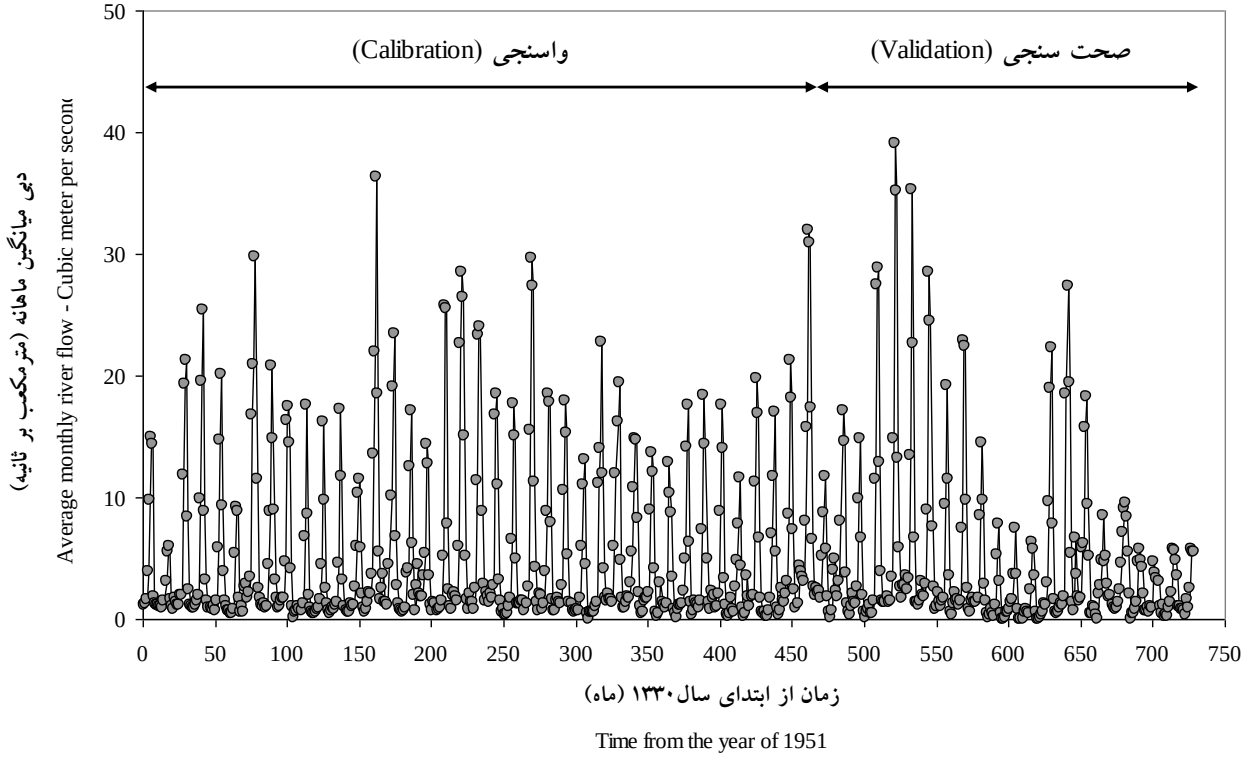
نمودار تغییرات میانگین ماهانه دبی در طی دوره آماری ۱۳۹۰-۱۳۳۰ در شکل ۴ و مشخصات آماری داده‌های مورد استفاده در جدول ۲ ارائه گردیده است.

برای پیش‌بینی دبی رودخانه شهرچای از داده‌های دبی مشاهداتی در زمان‌های ۱، ۲ و ۳ ماه قبلی (Q_{t-1}, Q_{t-2},Q_{t-3}) استفاده گردید. در نهایت دقت روش‌های مورد مطالعه با استفاده از پارامترهای آماری جذر میانگین مربعات خطا، میانگین خطای مطلق و ضریب همبستگی مورد بررسی قرار گرفت.

نتایج

به‌منظور بررسی امکان استفاده از توابع مختلف در روش برنامه-ریزی ژنتیک، سه حالت ترکیبی از توابع مورد اشاره در جدول ۱ مورد استفاده قرار گرفتند (جدول ۳). بعد از انجام محاسبات تشریح شده برای روش‌های برنامه‌ریزی ژنتیک و مدل درختی M5 در مرحله واسنجی، دقت روش‌های مذکور در مرحله صحت‌سنجی و بر اساس معیارهای آماری (روابط ۲ تا ۴) در جدول ۴ ارائه شده است.

همان‌طوری که از جدول ۴ مشاهده می‌گردد، در حالت محاسبه دبی رودخانه با پارامتر ورودی Q_{t-1}، مدل درختی M5 با داشتن جذر میانگین مربعات خطای ۴/۷۹۰۷ در مقایسه با بهترین روش برنامه‌ریزی ژنتیک (برنامه‌ریزی ژنتیک-۲) با جذر میانگین مربعات خطای ۴/۸۲۳۳ عملکرد بهتری از خود نشان داده است. همچنین مشاهده می‌گردد که افزایش توابع مورد استفاده برای روش برنامه-ریزی ژنتیک و پیچیده کردن فرمول تحلیلی استخراجی، تأثیر مثبتی در کاهش خطا نداشته است. برخلاف روند مشاهده شده در حالت قبل، در حالت محاسبه دبی رودخانه با پارامترهای ورودی Q_{t-1}, Q_{t-2}، روش برنامه‌ریزی ژنتیک-۲ با داشتن جذر میانگین مربعات خطای ۳/۳۵۰۱ در مقایسه با دیگر روش‌های برنامه‌ریزی ژنتیک و همچنین مدل درختی M5 با خطای ۳/۸۴۸۰ عملکرد مطلوب‌تری داشته است.



شکل ۴- نمودار تغییرات میانگین ماهانه دبی رودخانه شهرچای در طی دوره آماری ۱۳۹۰-۱۳۳۰

Figure 4- Fluctuations of monthly average river flow of Shaharchay river in time period of 1951-2011

جدول ۲- مشخصات آماری داده‌های رودخانه شهرچای

Table 2- Statistical parameters of Shaharchay river data

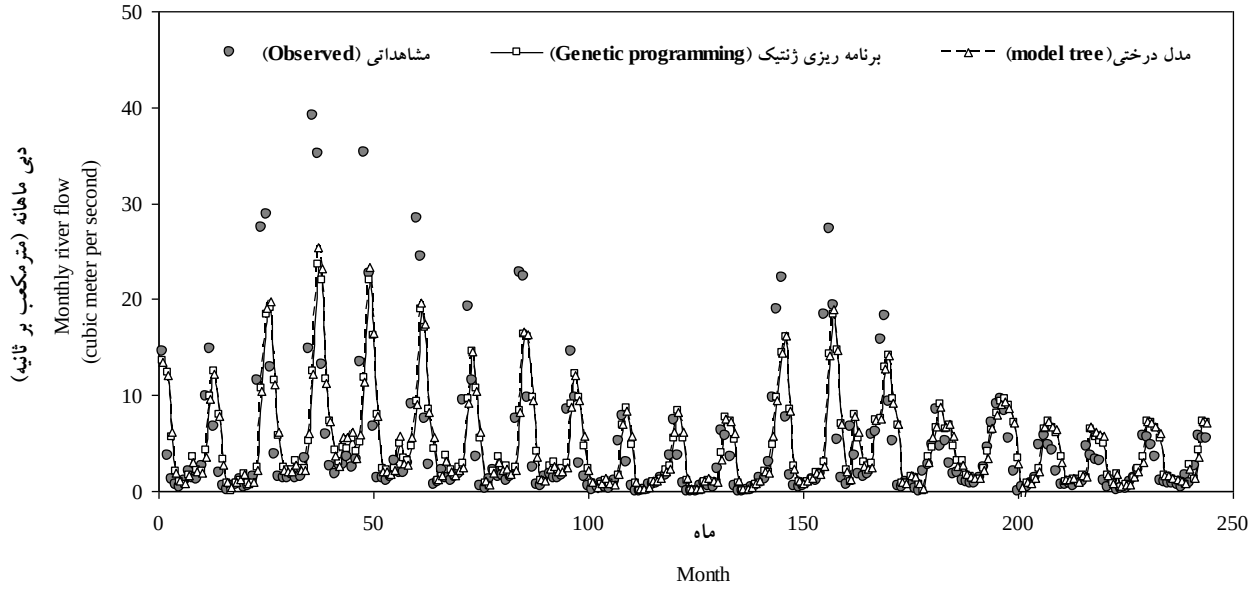
نوع داده‌ها Kind of data	Xmean (m^3/s)	Xmax (m^3/s)	Xmin (m^3/s)	Sx (m^3/s)	Csx (-)
داده‌های واسنجی Calibration data	5.019	39.090	0.000	6.721	2.104
داده‌های صحت‌سنجی Validation data	5.261	36.330	0.000	6.692	1.830
کل داده‌ها All data	4.537	39.090	0.000	6.767	2.660

Xmean: میانگین داده ها — Average of data
Xmax: حداکثر داده ها — Maximum of data
Csx: چولگی — Skewness
Xmin: حداقل داده ها — Minimum of data
Sx: انحراف معیار داده ها — Standard deviation

جدول ۳- توابع مورد استفاده در روش برنامه‌ریزی ژنتیک

Table 3- Functions which is used in genetic programming

روش Method	توابع مورد استفاده Used functions
برنامه‌ریزی ژنتیک – ۱ Genetic programming -1	+, -, ×, /, √
برنامه‌ریزی ژنتیک – ۲ Genetic programming -2	+, -, ×, /, √, Ln(x), ex, Power
برنامه‌ریزی ژنتیک – ۳ Genetic programming -3	+, -, ×, /, √, Ln(x), exp, Power, Sin, Cosine, Arctangent



شکل ۵- مقادیر مشاهداتی و پیش‌بینی شده دبی رودخانه شهرچای با مدل‌های برنامه‌ریزی ژنتیک-۲ و مدل درختی M5 و توسط پارامتر ورودی Q_{t-1}

Figure 5- Observed and predicted Shaharchay river flow by genetic programming-2 and M5 model tree using Q_{t-1}

جدول ۴- عملکرد مدل‌های مورد مطالعه در پیش‌بینی آبدهی رودخانه شهرچای در ایستگاه هیدرومتری بند

Table 4- Assessment of studied models in prediction of Shaharchay river flow in band hydrometric station

ورودی‌های مدل Inputs of the model	مدل Model	پارامترهای آماری Statistical parameters		
		RMSE (m^3/s)	MAE (m^3/s)	R
Q_{t-1}	برنامه‌ریزی ژنتیک – ۱ Genetic programming -1	4.8567	2.9138	0.7300
	برنامه‌ریزی ژنتیک – ۲ Genetic programming -2	4.8233	2.9536	0.7077
	برنامه‌ریزی ژنتیک – ۳ Genetic programming -3	4.9127	3.0255	0.6985
	مدل‌درختی M5 M5 model tree	4.7907	2.8884	0.7058
Q_{t-1}, Q_{t-2}	برنامه‌ریزی ژنتیک – ۱ Genetic programming -1	3.3544	2.0485	0.8730
	برنامه‌ریزی ژنتیک – ۲ Genetic programming -2	3.3501	2.1373	0.8731
	برنامه‌ریزی ژنتیک – ۳ Genetic programming -3	3.5644	2.2415	0.8554
	مدل‌درختی M5 M5 model tree	3.8480	3.3828	0.8232
Q_{t-1}, Q_{t-2}, Q_{t-3}	برنامه‌ریزی ژنتیک – ۱ Genetic programming -1	3.5715	2.1417	0.8546
	برنامه‌ریزی ژنتیک – ۲ Genetic programming -2	3.3094	2.0399	0.8769
	برنامه‌ریزی ژنتیک – ۳ Genetic programming -3	3.6836	2.3289	0.8493
	مدل‌درختی M5 M5 model tree	3.5514	2.0877	0.8562

در نهایت در حالت محاسبه دبی رودخانه با پارامترهای ورودی Q_{t-1},Q_{t-2},Q_{t-3}، روش برنامه‌ریزی ژنتیک-۲ با دارا بودن خطای ۳/۳۰۹۴ در مقایسه با روش‌های برنامه‌ریزی ژنتیک ۱ و ۳ و همچنین مدل درختی M5 با خطای ۳/۵۵۱۴ عملکرد بهتری ارائه کرده است. در نهایت می‌توان نتیجه‌گیری نمود که روش برنامه‌ریزی ژنتیک-۲ با در نظر گرفتن پارامترهای ورودی Q_{t-1},Q_{t-2},Q_{t-3}، با داشتن کمترین جذر میانگین مربعات خطا و بیشترین ضریب همبستگی بهترین عملکرد را در پیش‌بینی جریان رودخانه شهرچای داشته است. همچنین از جدول ۴ استنباط می‌شود که افزایش توابع مثلثاتی تأثیر مثبتی در کاهش خطای مدل نداشته است.

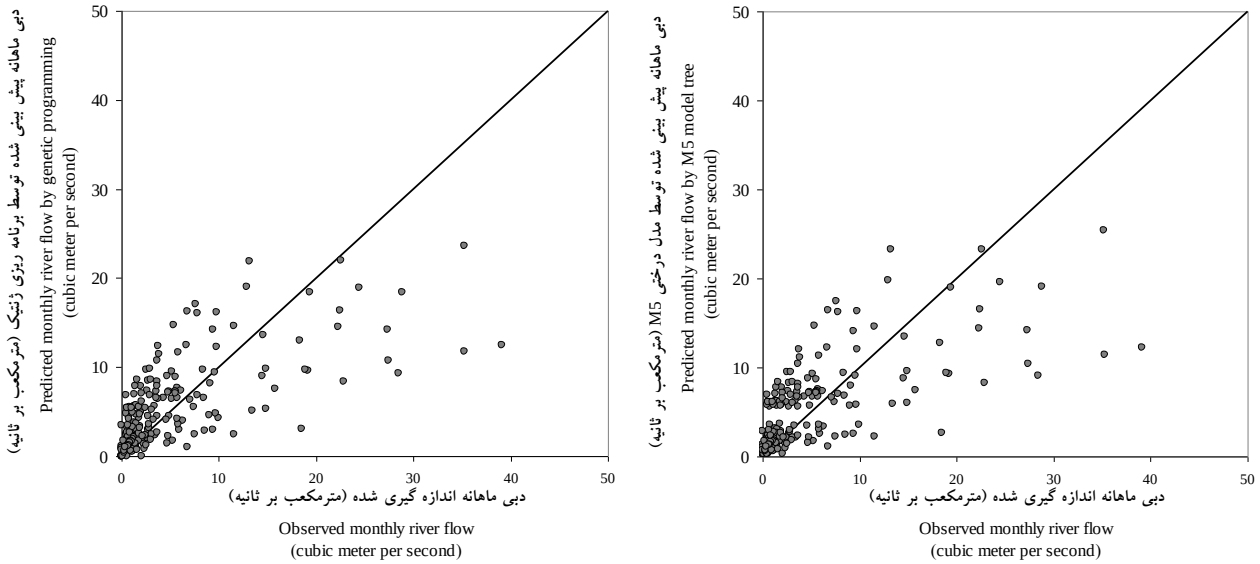
نمودارهای داده‌های شبیه‌سازی شده توسط بهترین روش برنامه‌ریزی ژنتیک (روش برنامه‌ریزی ژنتیک-۲) و مدل درختی M5 در مقابل داده‌های مشاهداتی در شکل‌های ۵ الی ۱۰ نشان داده شده است.

یکی از مزیت‌های روش برنامه‌ریزی ژنتیک نسبت به روش‌های دیگر هوش مصنوعی، ارائه فرمولی تحلیلی برای محاسبه پارامتر خروجی می‌باشد. پس از انجام محاسبات، روابط ارائه شده در جدول ۵ برای محاسبه دبی رودخانه شهرچای با داده‌های ورودی مختلف و روش‌های متفاوت برنامه‌ریزی ژنتیک حاصل شدند. لازم به ذکر است که آرگومان‌های توابع مثلثاتی ارائه شده در جدول ۵ بر حسب رادیان و مقادیر دبی بر حسب مترمکعب بر ثانیه می‌باشند.

همچنین روابط حاصل از مدل درختی M5 به‌صورت جدول ۶ برای محاسبه دبی رودخانه شهرچای با داده ورودی Q_{t-1}، جدول ۷ با داده‌های ورودی Q_{t-1}, Q_{t-2} و جدول ۸ با داده‌های ورودی Q_{t-1}, Q_{t-2},Q_{t-3} ارائه گردیده‌اند.

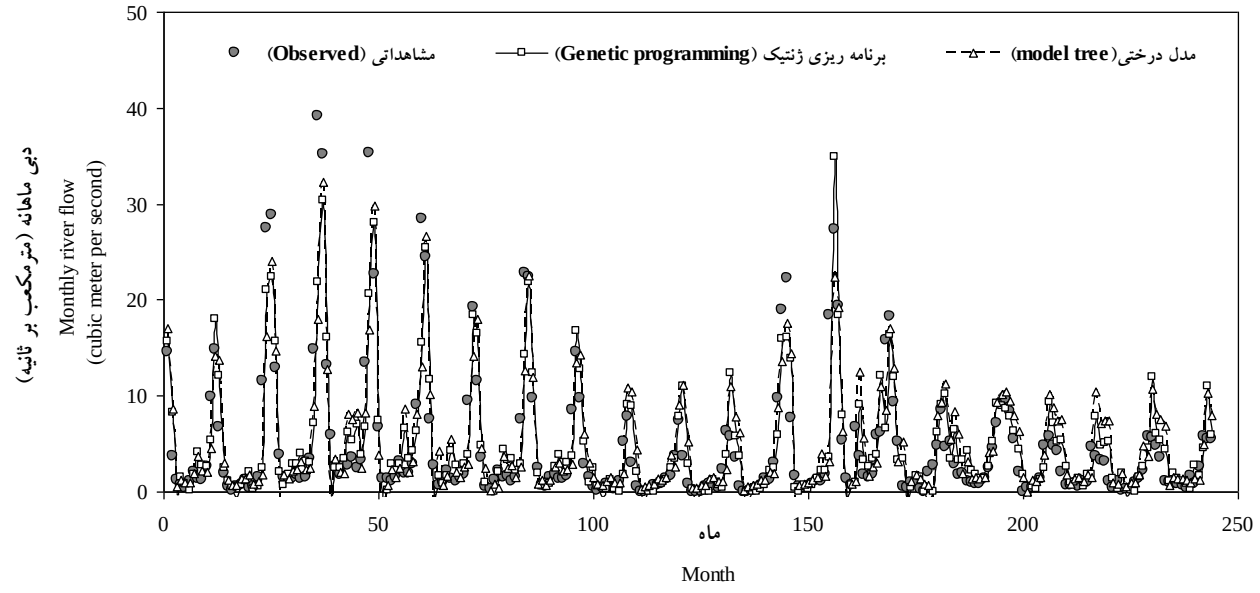
مقایسه نتایج حاصل از پژوهش حاضر با تحقیق انجام یافته توسط خلیلی و همکاران (۱۳) نشان داد که روش‌های بکار گرفته شده در پژوهش حاضر از دقت بالاتری برخوردار می‌باشند. محققین مذکور الگوی تلفیقی BL-ARCH را به‌منظور پیش‌بینی دبی رودخانه شهرچای ارومیه پیشنهاد نموده و اظهار داشتند که مدل تلفیقی

پیشنهادی با خطای ۴/۵۲، نسبت به مدل غیرخطی سری زمانی دو-خطی با خطای ۶/۷۷ عملکرد بهتری داشته است. اما بر اساس یافته‌های تحقیق حاضر، روش برنامه‌ریزی ژنتیک-۲ با دارا بودن خطای ۳/۳۰۹۴ و مدل درختی M5 با خطای ۳/۵۵۱۴ پیش‌بینی‌های دقیق‌تری را در مورد دبی‌های رودخانه شهرچای داشته‌اند. علاوه بر این، ظهیری و قربانی (۲۷) نیز دقت بالا و عملکرد قابل قبول مدل درختی M5 را در محاسبه دبی ماهانه چندین رودخانه با مقاطع مرکب گزارش نمودند.



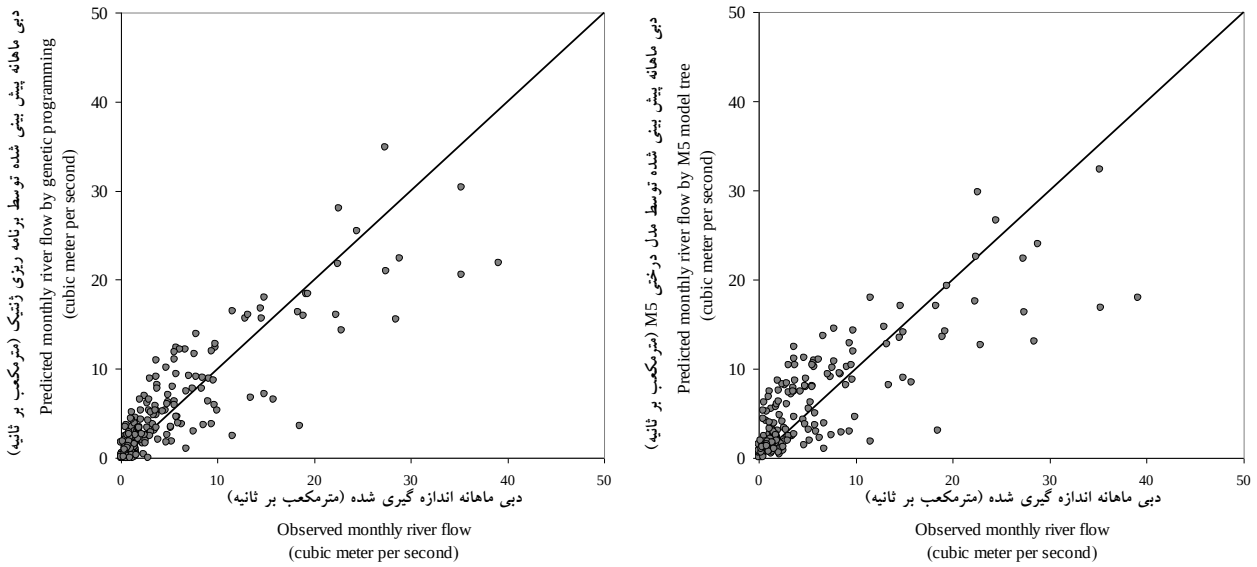
شکل ۶- نمودار پراکنش مقادیر مشاهداتی و پیش‌بینی شده دبی رودخانه شهرچای با مدل‌های برنامه‌ریزی ژنتیک-۲ و مدل درختی M5 و توسط پارامتر ورودی Q_{t-1}

Figure 6- Scatter plot of observed and predicted Shaharchay river flow by genetic programming-2 and M5 model tree using Q_{t-1}



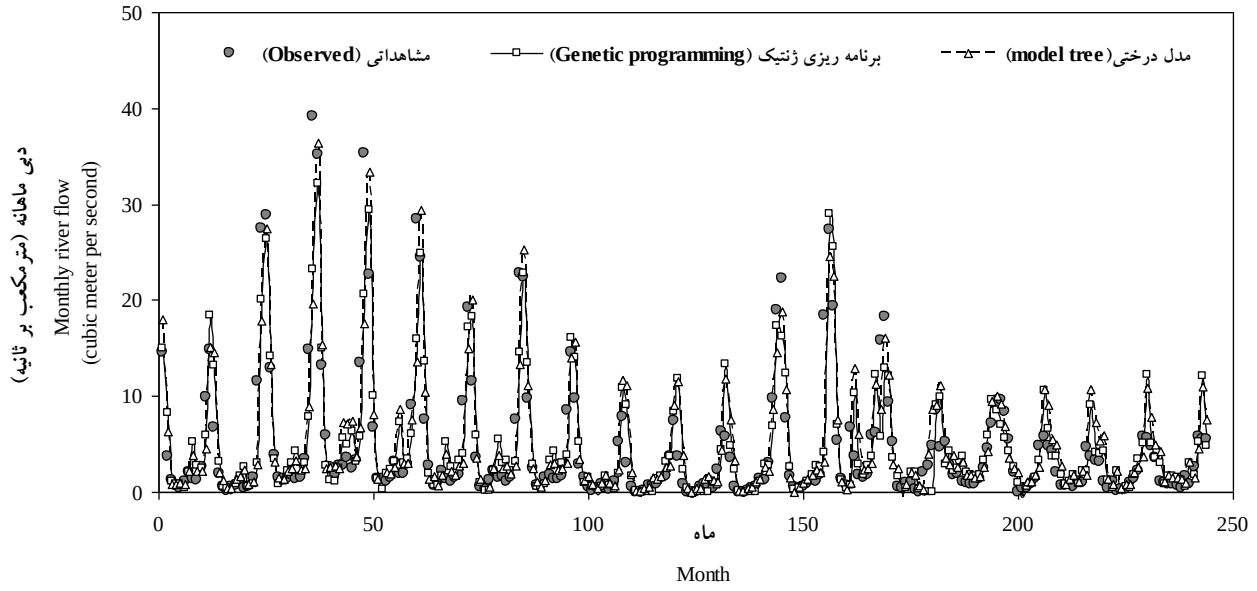
شکل ۷- مقادیر مشاهداتی و پیش‌بینی شده دبی رودخانه شهرچای با مدل‌های برنامه‌ریزی ژنتیک-۲ و مدل درختی M5 و توسط پارامترهای ورودی Q_{t-1}, Q_{t-2}

Figure 7- Observed and predicted Shaharchay river flow by genetic programming-2 and M5 model tree using Q_{t-1}, Q_{t-2}



شکل ۸- نمودار پراکنش مقادیر مشاهداتی و پیش‌بینی شده دبی رودخانه شهرچای با مدل‌های برنامه‌ریزی ژنتیک-۲ و مدل درختی M5 و توسط پارامترهای ورودی Q_{t-1}, Q_{t-2}

Figure 8- Scatter plot of observed and predicted Shaharchay river flow by genetic programming-2 and M5 model tree using Q_{t-1}, Q_{t-2}



شکل ۹- مقادیر مشاهداتی و پیش‌بینی شده دبی رودخانه شهرچای با مدل‌های برنامه‌ریزی ژنتیک-۲ و مدل درختی M5 و توسط پارامترهای ورودی $Q_{t-1}, Q_{t-2}, Q_{t-3}$

Figure 9- Observed and predicted Shaharchay river flow by genetic programming-2 and M5 model tree using $Q_{t-1}, Q_{t-2}, Q_{t-3}$

مقادیر R^2 و RMSE در تحقیق مذکور به ترتیب ۰/۹۳۸ و ۰/۴۴۶ محاسبه شد. در پژوهشی دیگر، داننده مهر و مجدزاده طباطبایی (۶) دقت روش‌های برنامه‌ریزی ژنتیک و شبکه‌های عصبی مصنوعی در پیش‌بینی جریان رودخانه آبسرده در استان لرستان را مورد مطالعه قرار داده و در نهایت روش برنامه‌ریزی ژنتیک را به‌عنوان یک روش صریح و دقیق برای پیش‌بینی جریان رودخانه‌ها پیشنهاد نمودند.

جدول ۵- روابط تحلیلی به‌دست‌آمده از روش‌های متفاوت برنامه‌ریزی ژنتیک برای پیش‌بینی دبی رودخانه شهرچای

Table 5- Resulted analytical formulations from different genetic programming models for prediction of Shaharchay river flow

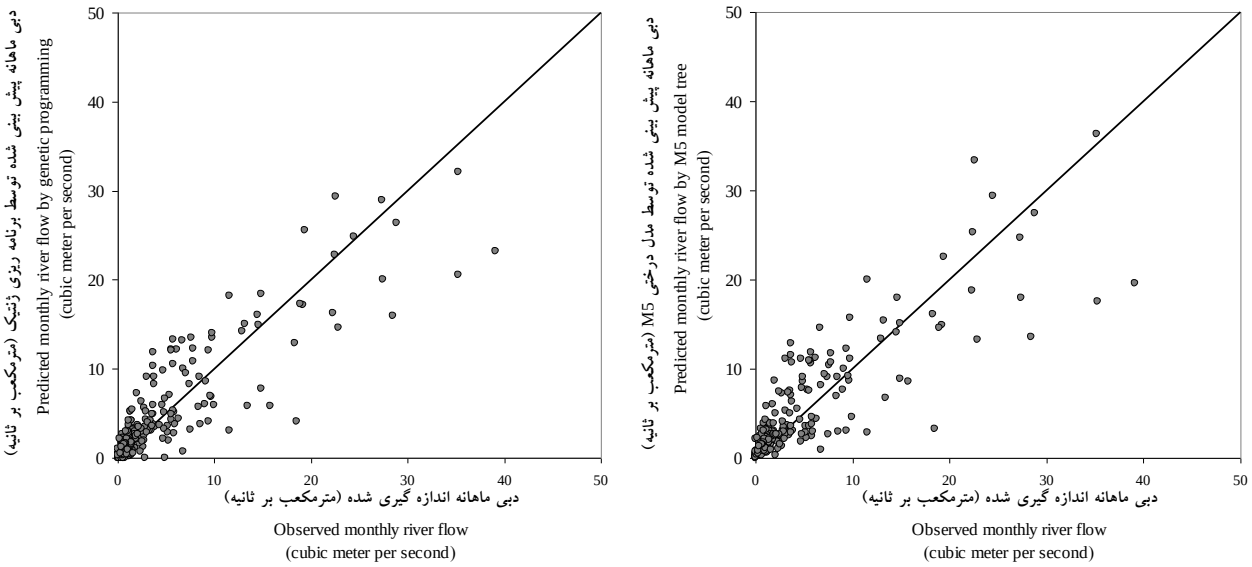
ورودی‌های مدل Inputs of the model	مدل Model	فرمول تحلیلی Analytical formula
Q_{t-1}	برنامه‌ریزی ژنتیک – ۱ Genetic programming -1	$Q_t = 1.16164 - \frac{1.28891}{15.4317 - Q_{t-1}} + 0.46432 \times (Q_{t-1} - 1.0069) + \frac{4.89395 \times Q_{t-1}}{Q_{t-1} + (7.4115/Q_{t-1})}$
	برنامه‌ریزی ژنتیک – ۲ Genetic programming -2	$Q_t = 2Q_{t-1} - 4.59897 + \exp(-Q_{t-1}) \times (0.470093 - Q_{t-1}) - \frac{1.18622 + Ln(Q_{t-1})}{1.18622 - Q_{t-1} + {Q_{t-1}}^2} + 1.60499(4.29288 - Q_{t-1} + Ln(Q_{t-1}))$
	برنامه‌ریزی ژنتیک – ۳ Genetic programming -3	$Q_t = 2.06184 \times (Q_{t-1})^{2/3} + \cos(1.75458 \times {Q_{t-1}}^2 + (9.29318 \times \sin(Q_{t-1}))/Q_{t-1}) + \cos(\arctan(5.58344 - Q_{t-1})^3 \times \cos(5.23111 - Q_{t-1})^3) - 0.113096 \times Q_{t-1} \times \cos(Q_{t-1})$
Q_{t-1}, Q_{t-2}	برنامه‌ریزی ژنتیک – ۱ Genetic programming -1	$Q_t = Q_{t-1} + \frac{3.08585 \times Q_{t-1}}{2.414 + Q_{t-2}} - 0.0480611 \times Q_{t-2} \times \left(0.41425 \times Q_{t-1} + \frac{8.61923}{Q_{t-2}}\right) + \frac{9.99295 - 3Q_{t-2}}{Q_{t-2} - 2.92712 + (9.99295/Q_{t-2})}$
	برنامه‌ریزی ژنتیک – ۲ Genetic programming -2	$Q_t = 2.24425 - (Q_{t-1} + 2.83551) \times \exp(-Q_{t-1}) - Q_{t-2} + 0.628713 \times (Q_{t-2} - Q_{t-1}) + +1.40657 \times Q_{t-1} + \frac{Q_{t-1}}{Q_{t-2} - 2.98038 + (3.87503/Q_{t-2})}$
	برنامه‌ریزی ژنتیک – ۳ Genetic programming -3	$Q_t = 1.7322 \times {Q_{t-1}}^{1/4} + 0.177459(Q_{t-2} - Q_{t-1}) + Q_{t-1} - 0.816559 \times Q_{t-2} + \arctan(Q_{t-1} - Q_{t-2})^2 \times e^{\cos((2.42914/Q_{t-1})} + \sin(\log({Q_{t-2}}^2))$
Q_{t-1}, Q_{t-2}, Q_{t-3}	برنامه‌ریزی ژنتیک – ۱ Genetic programming -1	$Q_t = \frac{8.16467 \times Q_{t-1} \times Q_{t-1}}{8.16467 \times Q_{t-1} + 9.95773 \times Q_{t-2} \times Q_{t-2}} - \frac{Q_{t-1}}{9.93353 - \frac{0.100669 \times Q_{t-3}}{3.72287 - Q_{t-3}}} + \frac{15.3807 \times Q_{t-1}}{5.57239 + Q_{t-1} + Q_{t-3} \times Q_{t-3}}$
	برنامه‌ریزی ژنتیک – ۲ Genetic programming -2	$Q_t = Q_{t-1} - \exp\left(-\frac{0.419237 \times \exp(Q_{t-1} - Q_{t-2})}{Q_{t-2}}\right) + \frac{8.20132 \times Q_{t-1} \times Q_{t-2}}{\exp(Q_{t-2}) + Q_{t-1} + Q_{t-3}} + \frac{Q_{t-1} \times Q_{t-3} - Q_{t-1} - Q_{t-3}}{-9.78882 - \frac{1}{(Q_{t-3})^{9.92569}}} - Q_{t-3}$
	برنامه‌ریزی ژنتیک – ۳ Genetic programming -3	$Q_t = \left((Q_{t-1} - \sin(0.765961 \times Q_{t-1}))/({Q_{t-2}}^{2/3})\right) + \left((Q_{t-1} - \sin(7.55762 - Q_{t-2}))/({Q_{t-3}}^{1/3})\right) + \sin\left(\sqrt{Q_{t-2}} - \sin(Q_{t-2} - 2 \times Q_{t-3})\right)$

آرگومان‌های توابع مثلثاتی ارائه شده بر حسب رادیان و مقادیر دبی بر حسب مترمکعب بر ثانیه می‌باشند.

جدول ۶- مدل‌های ریاضی به‌دست‌آمده از مدل درختی M5 برای پیش‌بینی شده دبی رودخانه توسط پارامتر ورودی Q_{t-1}

Table 6- Resulted mathematical models from M5 model tree for prediction of river flow using input parameter of Q_{t-1}

شرایط ورودی Input conditions	رابطه محاسبه دبی رودخانه Mathematical formula for computing river flow
Q_{t-1}	Q_t
$Q_{t-1} \leq 2.67$	$Q_t = 1.2674 \times Q_{t-1} + 0.3339$
$Q_{t-1} > 2.67$	$Q_t = 0.543 \times Q_{t-1} + 4.1532$



شکل ۱۰- نمودار پراکنش مقادیر مشاهداتی و پیش‌بینی شده دبی‌رودخانه شهرچای با مدل‌های برنامه‌ریزی ژنتیک-۲ و مدل درختی M5 و توسط پارامترهای ورودی $Q_{t-1}, Q_{t-2}, Q_{t-3}$

Figure 10- Scatter plot of observed and predicted Shaharchay river flow by genetic programming-2 and M5 model tree using $Q_{t-1}, Q_{t-2}, Q_{t-3}$

جدول ۷- مدل‌های ریاضی به دست آمده از مدل درختی M5 برای پیش‌بینی شده دبی رودخانه توسط پارامترهای ورودی Q_{t-1}, Q_{t-2}

Table 7- Resulted mathematical models from M5 model tree for prediction of river flow using input parameter of Q_{t-1}, Q_{t-2}

شرایط ورودی Input conditions		رابطه محاسبه دبی رودخانه Mathematical formula for computing river flow
Q_{t-1}	Q_{t-2}	Q_t
$Q_{t-1} \leq 1.175$	All values	$Q_t = 1.1401 \times Q_{t-1} - 0.3475 \times Q_{t-2} + 0.7538$
$1.175 < Q_{t-1} \leq 2.67$	$Q_{t-2} \leq 1.03$	$Q_t = 0.7029 \times Q_{t-1} + 6.2287 \times Q_{t-2} - 2.2456$
$1.175 < Q_{t-1} \leq 2.67$	$1.03 < Q_{t-2} \leq 2.42$	$Q_t = 1.892 \times Q_{t-1} - 0.1058 \times Q_{t-2} - 0.169$
$1.175 < Q_{t-1} \leq 2.67$	$Q_{t-2} > 2.42$	$Q_t = 1.0326 \times Q_{t-1} - 0.2073 \times Q_{t-2} + 0.6047$
$Q_{t-1} > 2.67$	All values	$Q_t = 0.9048 \times Q_{t-1} - 0.663 \times Q_{t-2} + 6.8348$

جدول ۸- مدل‌های ریاضی به‌دست‌آمده از مدل درختی M5 برای پیش‌بینی شده دبی رودخانه توسط پارامترهای ورودی $Q_{t-1}, Q_{t-2}, Q_{t-3}$

Table 8- Resulted mathematical models from M5 model tree for prediction of river flow using input parameter of $Q_{t-1}, Q_{t-2}, Q_{t-3}$

شرایط ورودی Input conditions			رابطه محاسبه دبی رودخانه Mathematical formula for computing river flow
Q_{t-1}	Q_{t-2}	Q_{t-3}	Q_t
$Q_{t-1} \leq 2.67$	All values	All values	$Q_t = 1.7424 \times Q_{t-1} - 0.2464 \times Q_{t-2} - 0.0619 \times Q_{t-3} + 0.4485$
$2.67 < Q_{t-1} \leq 14.635$	All values	$Q_{t-3} \leq 8.31$	$Q_t = 1.0535 \times Q_{t-1} - 0.6593 \times Q_{t-2} - 0.5963 \times Q_{t-3} + 7.0514$
$2.67 < Q_{t-1} \leq 14.635$	All values	$Q_{t-3} > 8.31$	$Q_t = 0.4704 \times Q_{t-1} - 0.1645 \times Q_{t-2} + 0.0266 \times Q_{t-3} + 1.0327$
$Q_{t-1} > 14.635$	All values	$Q_{t-3} > 8.31$	$Q_t = 0.6259 \times Q_{t-1} - 0.1727 \times Q_{t-2} + 0.0166 \times Q_{t-3} - 0.1082$

یافته‌های هر دو تحقیق فوق که حاکی از برتری روش‌های برنامه‌ریزی ژنتیک و مدل درختی M5 نسبت به روش‌های دیگر می‌باشد، تطابق کاملی با نتایج حاصل از تحقیق حاضر دارد. همچنین روابط ریاضی حاصل شده از روش برنامه‌ریزی ژنتیک در این تحقیق (روابط ارائه شده در جدول ۵) و مدل درختی M5 (جدول‌های ۶ الی ۸) نشانگر کاربردی بودن مدل‌های مذکور می‌باشند.

نتیجه‌گیری کلی

پیش‌بینی جریان رودخانه با توجه به اهمیت آن در تحلیل خشکسالی و سیلاب، طراحی تأسیسات آبی، آبگیری از رودخانه‌ها، برنامه‌ریزی و بهره‌برداری از مخازن سدها، کنترل فرسایش و رسوب رودخانه‌ها و غیره از دیرباز مورد توجه مهندسان آب بوده است. در این تحقیق، از مدل‌های هوشمند برنامه‌ریزی ژنتیک و مدل درختی M5 برای پیش‌بینی جریان ماهانه رودخانه شهرچای ارومیه واقع در استان آذربایجان غربی استفاده شده است. نتایج حاصل از این پژوهش نشان دادند که روش‌های برنامه‌ریزی ژنتیک-۲ و مدل درختی M5 در حالت‌های استفاده از حافظه‌های دبی یک، دو و سه ماه قبل با داشتن کمترین مقدار خطا، کارکرد مناسب‌تری در پیش‌بینی جریان رودخانه داشته‌اند. در نهایت روش برنامه‌ریزی ژنتیک در حالت استفاده از توابع ریاضی متشکل از چهار عملی اصلی، لگاریتم و توان و در نظر گرفتن پارامترهای ورودی Q_{t-1}, Q_{t-2},Q_{t-3} به‌عنوان روشی مناسب برای پیش‌بینی جریان ماهانه رودخانه شهرچای پیشنهاد گردید.

سپاسگزاری

بدین‌وسیله از داوران محترم به خاطر ارائه نظرات ارزشمندی که باعث ارتقای سطح علمی مقاله شده است، تشکر و قدردانی می‌گردد.

منابع

1- Alberg D., Last M., and Kandel A. 2012. Knowledge discovery in data streams with regression tree methods. Data Mining and Knowledge Discovery, 2(1): 69-78.
2- Alikhanzadeh A. 2013. Data mining. Olomrayaneh, Sari.
3- Aqil M., Kita I., Yano A., and Nishiygama A. 2005. A comparative study of artificial network and hourly behavior of run off. Journal of Hydrology, 337: 22-34.
4- Aytek A., and Alp M. 2008. An application of artificial intelligence for rainfall–runoff modeling. Journal of Earth System Science, 117(2): 145-155.
5- Bhattacharya B., and Solomatine D.P. 2006. Machine learning in sedimentation modeling. Neural Networks, 19(2): 208-214.
6- Danandehmehr A., and Majdzadeh Tabatabai M.R. 2010. Prediction of daily discharge trend of river flow based on genetic programming. Journal of Water and Soil, 24(2): 325-333. (in Persian with English abstract).
7- Ebrahimi Mohammadi S.H., and Boshri Se Ghaleh, M. 2011. Modeling and prediction of monthly discharge stream (case study: Qarasou River), 4th Iran Water Resources Management Conference, Amir kabir University of Technology, Tehran (In Persian).
8- El-Shafie A., RedaTaha M., and Noureldin A. 2007. A neuro-fuzzy model for inflow forecasting of the Nile river at Aswan high dam. Water Resource Management, 21: 533-556.
9- Fallahi M.R., Varvani H., and Goliyan S. 2012. Precipitation forecasting using regression tree model to flood control. 5th International watershed and water and soil resources management, 1-2 March, Kerman, Iran.
10- Farboudfam N., Ghorbani M.A., and Alami M.T. 2009. River flow prediction using genetic programming (Case study: Lighvan river watershed). Water and Soil Science, 19(1): 107-123. (in Persian with English abstract).
11- Firat M., and Gungor M. 2006. River flow estimation using adaptive neuro-fuzzy inference system. Journal of Mathematics and Computers in Simulation, 75(3-4): 87-96.
12- Ghobadian R., Ghorbani M.A., and Khalaj M. 2013. Comparison of performance of dynamic wave and gen expression programming methods to river flood routing. Journal of Water and Soil, 27(3): 592-602. (in Persian with English abstract).
13- Khalili K., Fakheri Fard A., Dinpaghoh Y., Ahmadi F., and Behmanesh J. 2013. Introducing and application of combined BL-ARCH model for daily river flow forecasting (Case study: Shahar-Chai river). Journal of Water and Soil, 27(2): 342-350. (in Persian with English abstract).
14- Khatibi R., Ghorbani M.A., Hasanpourkashani M., and Kisi O. 2010. Comparison of three artificial intelligence techniques for discharge routing. Journal of hydrology, 403(3-4): 201-212.
15- Khazaei m., and Mirzaei M.R. 2013. Comparison of artificial neural network and time series in prediction of monthly river flows. Journal of Watershed Engineering and Management, 5(2): 74-84. (in Persian).
16- Khu S.T., Liong S.Y., Babovic V., Madsen H., and Muttil N. 2001. Genetic programming and its application in real- time runoff forming. Journal of the American Water Resources Association, 37(2): 439-451.
17- Koza J.R. 1992. Genetic Programming: On the Programming of Computers by Means of Natural Selection. MIT Press.
18- Liu W.C., and Chen W.B. 2012. Prediction of water temperature in a subtropical subalpine lake using an artificial neural network and three-dimensional circulation models. Computers Geosciences, 45: 13-25.
19- Londhe S.N., and Dixit P.R. 2011. Forecasting Stream Flow Using Model Trees. International Journal of Earth

Sciences and Engineering, 4(6): 282-285.
20- Moradizadeh Kermani F., Ghorbani M.A., Dinpashoh Y., and Farsadizadeh D. Afshari H.R. 2013. Predicting model of river streamflow based on chaotic phase space reconstruction. Water and Soil Science, 22(4): 1-16. (in Persian with English abstract).
21- Nabizadeh M., Mosaedi A., Hesam M., Dehghani A.A., Zakerinia M., and Meftah, M. 2012. River flow forecasting using fuzzy inference system (FIS) and adaptive neuro-fuzzy inference system (ANFIS). Iran-Watershed Management Science & Engineering, 5(17): 7-14. (in Persian with English abstract).
22- 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). Journal of Water and Soil, 26(5): 1299-1307. (in Persian with English abstract).
23- Nourani, V., and Salehi, K. 2008. Rainfall-runoff modeling using Adaptive Neuro-Fuzzy network in comparison with Neural Network and Fuzzy Inference methods. CD's of 4th national congress of civil engineering, Tehran University, 8p. (In Persian).
24- Pal M. 2006. M5 model tree for land cover classification. International Journal of Remote Sensing, 27(4): 825-831.
25- Quinlan J.R. 1992. Learning with continuous classes. In proceedings AI, 90 (Adams & Sterling, Eds), Singapore.
26- Sattari M.T., Pal M., Apaydin H., and Ozturk F. 2013. M5 model tree application in daily river flow forecasting in Sohu stream, Turkey. Water Resources, 40(3): 233-242.
27- Zahiri A.R., and Ghorbani Kh. 2013. Flow discharge prediction in compound channels by using decision model tree M5. Journal of Water and Soil Conservation, 24(3): 113-132. (in Persian with English abstract).

Journal of Water and Soil
Vol. 29, No. 5, Nov.-Dec. 2015, p. 1190-1206

نشریه آب و خاک (علوم و صنایع کشاورزی)
جلد ۲۹، شماره ۵، آذر – دی ۱۳۹۴، ص. ۱۲۰۶-۱۱۹۰

Forecasting Shaharchay River Flow in Lake Urmia Basin using Genetic Programming and M5 Model Tree

S. Samadianfard[1*] - R. Delirhasannia

Received: 12-04-2014
Accepted: 19-04-2015

Introduction: Precise prediction of river flows is the key factor for proper planning and management of water resources. Thus, obtaining the reliable methods for predicting river flows has great importance in water resource engineering. In the recent years, applications of intelligent methods such as artificial neural networks, fuzzy systems and genetic programming in water science and engineering have been grown extensively. These mentioned methods are able to model nonlinear process of river flows without any need to geometric properties. A huge number of studies have been reported in the field of using intelligent methods in water resource engineering. For example, Noorani and Salehi (23) presented a model for predicting runoff in Lighvan basin using adaptive neuro-fuzzy network and compared the performance of it with neural network and fuzzy inference methods in east Azerbaijan, Iran. Nabizadeh et al. (21) used fuzzy inference system and adaptive neuro-fuzzy inference system in order to predict river flow in Lighvan river. Khalili et al. (13) proposed a BL-ARCH method for prediction of flows in Shaharchay River in Urmia. Khu et al. (16) used genetic programming for runoff prediction in Orgeval catchment in France. Firat and Gungor (11) evaluated the fuzzy-neural model for predicting Mendes river flow in Turkey. The goal of present study is comparing the performance of genetic programming and M5 model trees for prediction of Shaharchay river flow in the basin of Lake Urmia and obtaining a comprehensive insight of their abilities.

Materials and Methods: Shaharchay river as a main source of providing drinking water of Urmia city and agricultural needs of surrounding lands and finally one of the main input sources of Lake Urmia is quite important in the region. For obtaining the predetermined goals of present study, average monthly flows of Shaharchay River in Band hydrometric station has been gathered from 1951 to 2011. Then, two third of mentioned data were used for calibration and the rest were used for validation of study models including genetic programming and M5 model trees. It should be noted that for prediction of Shaharchay river flows, previous data of mentioned river in 1, 2 and 3 months ago (Q, Q, Q) were used.

Genetic programming: was first proposed by Koza (17). It is a generalization of genetic algorithms. The fundamental difference between genetic programming and genetic algorithm is due to the nature of the individuals. In genetic algorithm, the individuals are linear strings of fixed length (chromosomes). In genetic programming, the individuals are nonlinear entities of different sizes and shapes (parse trees). Genetic programming applies genetic algorithms to a "population" of programs, typically encoded as tree-structures. Trial programs are evaluated against a "fitness function". Then the best solutions are selected for modification and re-evaluation. This modification-evaluation cycle is repeated until a "correct" program is produced.

Model trees generalize the concepts of regression trees, which have constant values at their leaves. So, they are analogous to piece-wise linear functions. M5 model tree is a binary decision tree having linear regression function at the terminal nodes, which can predict continuous numerical attributes. Tree-based models are constructed by a divide-and-conquer method.

Results and Discussion: In order to investigate the probability of using different mathematical functions in genetic programming method, three combinations of the functions were used in the current study. The results showed that in the case of predicting river flows with Q, M5 model trees with root mean squared error of 4.7907 in comparison with genetic programming by the best mathematical functions and root mean squared error of 4.8233 had better performances. Obtained results indicated that adding more mathematical functions to the genetic programming and producing more complicated analytical formulations did not have positive effect in reducing prediction error. Unlike the previous observed trend, in case of predicting river flows with Q Q, the genetic programming method with root mean squared error of 3.3501 in comparison with M5 model trees with error of 3.8480 had more satisfied performance. Finally, in the case of predicting river flows with Q, Q,Q, the genetic programming method with root mean squared error of 3.3094 in comparison with M5 model trees with error of 3.5514 presented better predictions. As a result, it can be stated that genetic programming by the best mathematical functions and considering the input parameters of Q,Q,Q, by resulting less root mean squared error and high correlation coefficients had the best performances among others. Also, the results showed that adding

1,2- Assistant Professors, Department of Water Engineering, Faculty of Agriculture, University of Tabriz
(*-Corresponding Author Email: s.samadian@tabrizu.ac.ir)

more trigonometric functions did not improve the precisions of the predictions.

Conclusion: In this research, the intelligent models such as genetic programming and M5 model trees have been used for prediction of monthly flows of Shaharchay River located in East Azerbaijan, Iran. The obtained results showed that the genetic programming by the best mathematical functions and M5 model trees in case of considering the input parameters of Q,Q,Q, by less root mean squared error had the best performances in river flow predictions. As a conclusion, the genetic programming method by specific mathematical functions including four basic operations, logarithm, power and using input parameters of Q,Q,Q, has been proposed as the best and precise model for predicting Shaharchay River flows.

Keywords: Estimation, Flow discharge, Intelligence methods, Statistical parameters